BUSINESS INTELLIGENCE LABORATORY

Practice on a Classification Problem

Dataset

- ee_dataset.arff

- A dataset of 7.500 customers of a German electric power company

- Some customers intend to cancel their subscription (attribute canceler)

- A special promotion consisting of a discount on the price of electricity must be planned to prevent cancelers to abandon.

Data Science Lab

Task 1: Preprocessing

- ☐ Split the dataset into training and test
- ☐ Investigate the meaning of attributes from the provided documentation
- ☐ Study the distribution of data and the relevance of attributes
- ☐ If needed, create derived attributes

Task2: Classification

- Address the classification problem maximizing the following gain function instead of accuracy:

	No offer sent	Offer sent
Non Canceler	72,00 Euro	66,30 Euro
Canceler	0 Euro	43,80 Euro

Data Science Lab

Task 3: Lift Chart

☐ Assume to have limited amount of resources, so that at most 250 offers can be sent out. How many cancelers does your classifier can reach?

Task 4: Validation set

☐ Answer to the tasks using as test set a totally new set of data (ee_validation.arff). How do the performances of your classifier change?