Report I

Anna Szymanek (230042), Patryk Wielopolski (234891)

Introduction

Task descripion

Why was the study undertaken? What was the research question, the tested hypothesis or the purpose of the research?

- Problem description and research question formulation Short information on the specicity of the problem considered. What questions do we want to answer analysing the data? What potential benefits may result from the analysis? For example, the benefit could be: a better diagnostic method, better efficiency in detecting bad/goo d customers applying for a loan, separating groups of customers who can be targeted witha special offer, identifying relevant features/variables, etc. # End of the task description

Used data set [1].

Methods

Task description

When, where, and how was the study done? What materials were used or who was included in the study groups (patients, etc.)

- Methods and algorithms used in the project What methods / algorithms have been used? Which tasks these methods / algorithms were used for? (e.g. preliminary analysis and visualization, classifiaction, prediction, cluster analysis, etc.)
- Classification ** Objective - classification rule ** Methods: LDA, QDA, KNN, other ** Assessment of accuracy: Holdout / Cross Validation # End of the task description

Data analysis

Task description

- Data characteristics Data size, number of cases and features, types of features, information about missing values, information on unusual values (e.g. non-standard coding of missing values, etc.)
- Descriptive analysis and data visualization ** Objective *** Basic properties of variables / features (range, properties, distribution) *** Identification of missing values and outliers *** Analysis of correlation of features *** Initial assessment of discriminative ability of consecutive features (i.e. ability to separate objects from different classes) # End of the task description

Variable name	Type	Description
Bank client data		
Age	Numeric	Age of client
Job	Categorical	Type of job
Marital	Categorical	Marital status of client
Education	Categorical	Education status of client
Default	Categorical	Has credit in default?
Housing	Categorical	Has housing loan?
Loan	Categorical	Has personal loan?
Variables related with the last contact of the current campaign		
Contact	Categorical	Contact communication type
Month	Categorical	Last contact month of year
Day of week	Categorical	Last contact day of the week
Duration	Numeric	Last contact duration, in seconds
Other attributes		
Campaign	Categorical	Number of contacts performed during this campaign
Pdays	Numeric	Number of days that passed by after the client was last contacted from a previous campaign
Previous	Numeric	Nnumber of contacts performed before this campaign
Poutcome	Categorical	Outcome of the previous marketing campaign
Social and economic context attributes		
Emp.var.rate	Numeric	Employment variation rate - quarterly indicator
Cons.price.idx	Numeric	Consumer price index - monthly indicator
Cons.conf.idx	Numeric	Consumer confidence index - monthly indicator
Euribor3m	Numeric	Euribor 3 month rate - daily indicator
Nr. employed	Numeric	Number of employees - quarterly indicator
Outcome variable		
y	Categorical	Has the client subscribed a term deposit?

Table 1: Input variables

```
## # A tibble: 1 x 9
##    rows columns discrete_columns continuous_colu~ all_missing_col~
##   <int>   <int>            <int>            <int>            <int>
## 1 41188      21               11               10                0
## # ... with 4 more variables: total_missing_values <int>, complete_rows <int>,
## #   total_observations <int>, memory_usage <dbl>
```

Input variables 1. There is 41188 rows.

Duration - Important note: this attribute highly affects the output target (e.g., if duration=0 then y="no"). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

Pdays - 999 means client was not previously contacted

Classification

```
## INFO  [16:02:46.532] Benchmark with 9 resampling iterations
## INFO  [16:02:46.680] Applying learner 'classif.qda' on task 'full' (iter 3/3)
## INFO  [16:02:46.803] Applying learner 'classif.qda' on task 'full' (iter 1/3)
## INFO  [16:02:47.016] Applying learner 'classif.lda' on task 'full' (iter 1/3)
## INFO  [16:02:47.123] Applying learner 'classif.lda' on task 'full' (iter 2/3)
## INFO  [16:02:47.221] Applying learner 'classif.kknn' on task 'full' (iter 2/3)
## INFO  [16:02:54.156] Applying learner 'classif.kknn' on task 'full' (iter 3/3)
## INFO  [16:03:01.365] Applying learner 'classif.qda' on task 'full' (iter 2/3)
## INFO  [16:03:01.451] Applying learner 'classif.kknn' on task 'full' (iter 1/3)
## INFO  [16:03:08.471] Applying learner 'classif.lda' on task 'full' (iter 3/3)
## INFO  [16:03:08.605] Finished benchmark
```

Results

Task description

What answer was found to the research question; what did the study find? Was the tested hypothesis true?

Results presented in the form of corresponding tables, graphs and diagrams. Note that only the most important results should be included in the report, whereas additional results can be added as attachments.

** Mehtods *** Summary statistics *** Plots # End of the task description

Discussion

Task description

What might the answer imply and why does it matter? How does it fit in with what other researchers have found? What are the perspectives for future research?

- Conclusions Precise conclusions: what can be concluded from the analyses carried out? How these conclusions could be put into practice? (e.g. development of a new / better strategy in the company, new / better diagnostic methods, etc.)
- Further research suggestions Short information on further possible directions of research (what could / should be further studied and what additional methods / algorithms could be used?) # End of the task description

References

[1] P. Cortez S. Moro and P. Rita. A data-driven approach to predict the success of bank telemarketin. *Decision Support Systems*, 2014.