University of Padova

Master Degree in *Computer Science*

a.y. 2017/2018

Data Mining

Teacher: Annamaria Guolo

Practical assessment: June 20, 2018

INSTRUCTIONS: The examination takes 2 hours and 30 minutes. Write a brief report explaining the data analysis performed and the results. Provide a printed copy of the report. Remember to write your i) name and surname, ii) enrolment number, iii) date. Reports without the last three information will not be evaluated.

In the following you can find a description of the data and a brief guide for the analysis to be carried out. The use of the material of the course (slides, notes) is allowed. Internet is not allowed.

Dataset `icua`: data refer to patients admitted to an intensive care unit

- `sex`: Patient's sex (Female/Male)
- `race`: Patient's race (White, Black, Other)
- `service`: Type of service (Medical/ Surgical)
- `cancer`: Is cancer involved? No/Yes
- `renal`: Is chronic renal failure involved? No/Yes
- `infection`: Is infection involved? No/Yes
- `cpr`: Did the patient get CPR (CardioPulmonary Resuscitation) prior to admission? No/Yes
- `previous`: Previous admission to ICU within 6 months? No/Yes
- `pO2low`: Partial oxygen level from blood gases under 60? No/Yes
- `pHlow`: pH from blood gas under 7.25? No/Yes
- `pCO2hi`: Partial carbon dioxide level from blood gas over 45? No/Yes
- `bicarbonateLow`: Bicarbonate from blood gas under 18? No/Yes
- `creatinineHi`: Creatinine from blood gas over 2.0? No/Yes
- `status`: Patient's status (1 Died, 0 Lived)
- `age`: Patient's age (in years)
- `conscious`: consciousness (Yes), coma (No)

The aim of the analysis is the evaluation of the association of the patient's status to the other variables.

1. Consider the dataset composed by `status`, `age`, `conscious`, `service`, `previous`. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from `R` and the graphical evaluation fo the model/models that are considered most useful in order to explain the analysis and the results.

2. Consider all the variables in the dataset. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from `R` and the graphical evaluation fo the model/models that are considered most useful in order to explain the analysis and the results.

If needed, please report the seed used in your analyses.