**CMPT459 Fall 2018**

**Data Mining**

**Martin Ester**

**TA: Sahand Khakabimamaghani**

**Comments on Programming Assignment 2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Code** | **Max. Penalty** | **Description** |
| **1** | 1a | 4 | Sampled twice for train and test indexes |
|  | 1b | 2 | Sampled 75% as the train set! |
|  | 1c | 6 | No code for splitting |
|  |  |  |  |
|  |  |  |  |
| **2** | 2a | 3 | Missing values with nchar==0 not reported |
|  | 2b | 5 | Not reported missing values per feature (overall report only) |
|  |  |  |  |
|  |  |  |  |
| **3** | 3a | 5 | Selected 'survived', 'boat', or 'body', which are not available at the time of accident. |
|  | 3b | 5 | Not selected some of the potentially relevant features 'pcalss', 'sex', 'age', 'sibsb', 'parch', and 'fare', which are available at the time of accident, without any convincing explanation. |
|  |  |  |  |
|  |  |  |  |
| **4** | 4a | 4 | Removed some of the records with missing features (only should remove records with missing class) |
|  | 4b | 2 | Did not consider the class label ('survived') when taking mean/median of 'age' or 'fare' |
|  | 4c | 10 | Removing the records as the only solution (should have followed slide 48) |
|  | 4d | 5 | No changes applied to the data |
|  | 4e | 5 | Used mode for continuous features |
|  | 4f | 5 | Used median/mean for categorical features |
|  | 4g | 2 | Wrong average for 'age' or 'fare'! |
|  | 4h | 5 | Added new features |
|  | 4i | 4 | Mismatch between report and code |
|  | 4j | 5 | Incomplete predictive model for missing value imputation |
|  | 4k | 5 | Using maximum value instead of the most frequent |
|  | 4l | 5 | Bug in code (interpreter issued error) |
|  | 4m | 20 | No code! |
|  |  |  |  |
|  |  |  |  |
| **5** | 5a | 3 | Reported training accuracy instead of test accuracy |
|  | 5b | 2 | Reported wrong accuracy (miscalculations) |
|  | 5c | 1 | Reported wrong accuracy (mismatch with code output) |
|  | 5d | 2 | Did not used the cleaned data for training |
|  | 5e | 10 | Did not compute the test accuracy |
|  | 5f | 2 | Wrong or no tree size reported |
|  | 5g | 2 | Unsucessful tree training due to features with more than 32 levels |
|  | 5h | 2 | Included 'boat' or 'body' in the classifier |
|  | 5i | 4 | Retrained on test data |
|  | 5j | 5 | Bug in code (interpreter issued error) |
|  |  |  |  |
|  |  |  |  |
| **6** | 6a | 3 | Not answered "What knowledge about the survival of passengers can you learn from the decision tree?" |
|  | 6b | 15 | No or incomplete answer provided |
|  | 6c | 8 | Not used the tree structure |
|  | 6d | 9 | Not mentioned some of the important features ('sex', 'pclass', 'age') |
|  | 6e | 8 | Wrong analysis |
|  | 6f | 5 | Bug in code (interpreter issued error) |
|  |  |  |  |
|  |  |  |  |
| **7** | 7a | 5 | Wrong or no size reported for pruned tree |
|  | 7b | 15 | No answer provided |
|  | 7c | 8 | Did not use cross-validation |
|  | 7d | 5 | Bug in code (interpreter issued error) |
|  |  |  |  |
|  |  |  |  |
| \* |  | 50 | Not sent the .Rmd file |