



Evaluation Exercise IV

Bergner, Borchert, da Cruz, Konak, Dr. Schapranow

Data Management for Digital Health

Winter 2019

Exercise IV

Topics

Sensor Data

Unsupervised Learning

Medical Use Case Nephrology

Digital Health Data Analysis Tools

**Evaluation Exercise
IV**

Data Management for
Digital Health, Winter
2019

2

Exercise IV

Key Stats

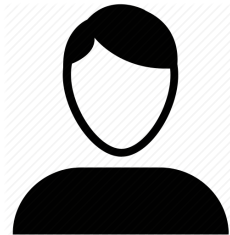
25 Questions
50 Points

13 Students
13 Passed

Average score
42.4 / 84.8 %

Average time
95.8 min

<< 3h



Evaluation Exercise
IV

Data Management for
Digital Health, Winter
2019

3

Q8. What can unsupervised learning be used for?

Find clusters of the data

Low-dimensional representations of the data

Predicting a future response in a time series with a regression line

Classification

**Evaluation Exercise
IV**

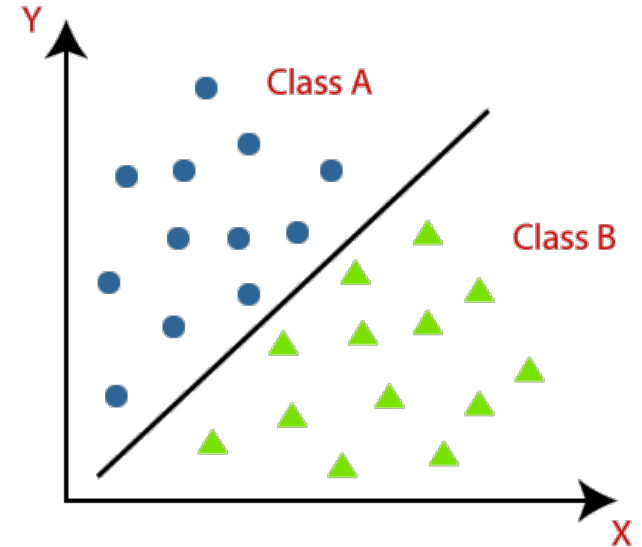
Data Management for
Digital Health, Winter
2019

4

Q8. What can unsupervised learning be used for?

In machine learning and statistics, **classification** is the problem of identifying to which of a set of categories (sub-populations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known. ~Wikipedia

The **Classification** algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. ~ JavaPoint



Evaluation Exercise
IV

Data Management for
Digital Health, Winter
2019

5

Q8. What can unsupervised learning be used for?

Find clusters of the data



Low-dimensional representations of the data



Predicting a future response in a time series with a regression line



Classification



**Evaluation Exercise
IV**

Data Management for
Digital Health, Winter
2019

6

Q14. On which gene is the given variant located? Consider possible gene aliases. Hint: for gene aliases, check GeneCards.

NR1C3

PPARG

IL7

SMMCI

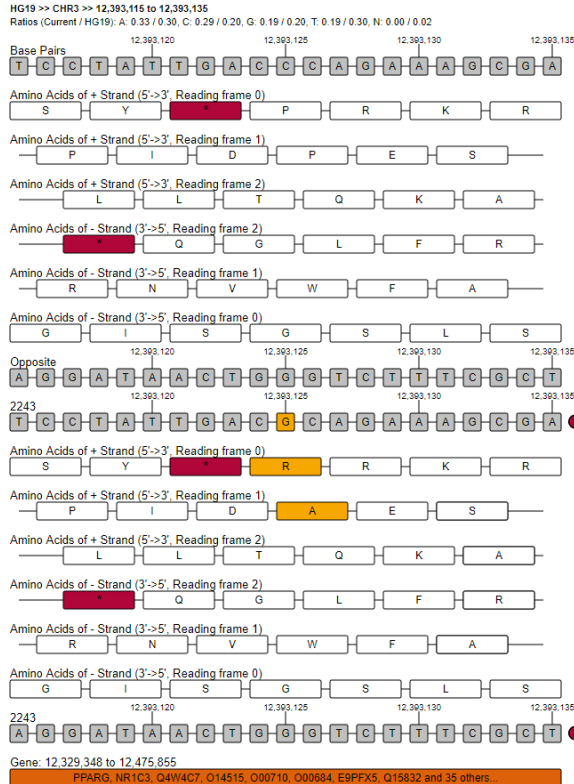
**Evaluation Exercise
IV**

Data Management for
Digital Health, Winter
2019

7

Q14. On which gene is the given variant located? Consider possible gene aliases. Hint: for gene aliases, check GeneCards.

Tasks » 2243 » Chromosome 3 » Position 12393125



Aliases for PPARG Gene

Aliases for PPARG Gene

- Peroxisome Proliferator Activated Receptor Gamma ^{2 3 5}
- Peroxisome Proliferator-Activated Receptor Gamma ^{2 3 4}
- Nuclear Receptor Subfamily 1 Group C Member 3 ^{3 4}
- PPAR-Gamma ^{3 4}
- NR1C3 ^{3 4}
- Peroxisome Proliferator-Activated Nuclear Receptor Gamma Variant 1 ³

- Peroxisome Proliferative Activated Receptor, Gamma ²
- PPARGgamma ³
- PPARG1 ³
- PPARG2 ³
- CIMT1 ³
- GLM1 ³

External Ids for PPARG Gene

HGNC: 9236 Entrez Gene: 5468 Ensembl: ENSG00000132170 OMIM: 601487 UniProtKB: P37231

Previous GeneCards Identifiers for PPARG Gene

GC03P012258, GC03P012304, GC03P012305, GC03P012263, GC03P012328

Search aliases for PPARG gene in PubMed and other databases

Evaluation Exercise IV

Data Management for Digital Health, Winter 2019

Q14. On which gene is the given variant located? Consider possible gene aliases. Hint: for gene aliases, check GeneCards.

NR1C3



PPARG



IL7



SMMCI

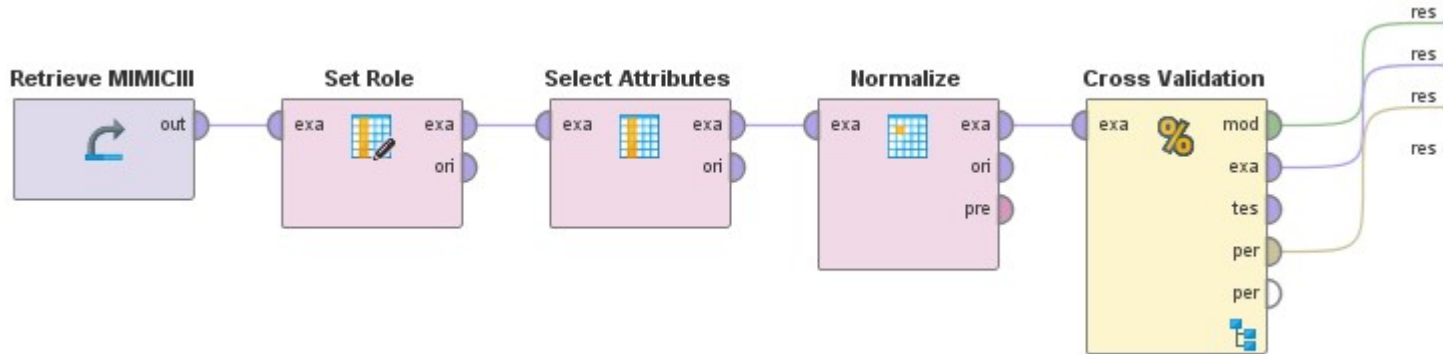


**Evaluation Exercise
IV**

Data Management for
Digital Health, Winter
2019

Q23. Sometimes, a useful strategy to achieve better performance in a clinical model [...]

[...] is to normalize numerical attributes. Modify your clinical model developed previously to include the Normalize operator. To evaluate performance, use the Cross Validation operator (k=10) as laid out below. Please run the model and choose the correct alternatives:



Evaluation Exercise IV

Data Management for
Digital Health, Winter
2019
10

Q23. Sometimes, a useful strategy to achieve better performance in a clinical model [...]

The performance achieved remained qualitatively unchanged as compared to the previous model

Decision trees usually do not require normalization of numerical attributes, what may explain the performance of this model being similar to the previous one

The cross-validation accuracy presents a very high standard deviation and is therefore unreliable

Class precision for AKI=yes is now substantially better than in the previous model

Q23. Sometimes, a useful strategy to achieve better performance in a clinical model [...]

The performance achieved remained qualitatively unchanged as compared to the previous model



Decision trees usually do not require normalization of numerical attributes, what may explain the performance of this model being similar to the previous one



The cross-validation accuracy presents a very high standard deviation and is therefore unreliable



Class precision for AKI=yes is now substantially better than in the previous model



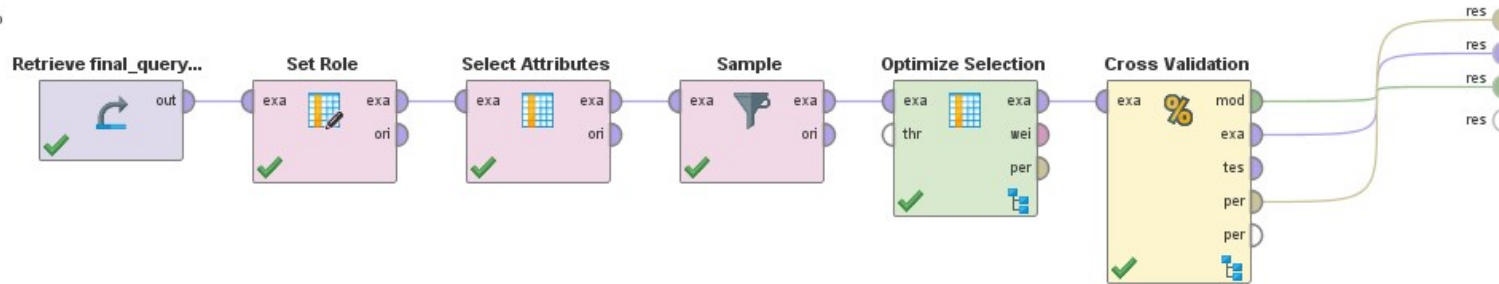
Evaluation Exercise
IV

Data Management for
Digital Health, Winter
2019

12

Q25. Feature selection is an important step in the development of a clinical prediction model. [...]

[...] It may simplify the model, improving interpretability, as well as make it easier for deployment in production. Often expert knowledge comes into play, but there are also strategies for automatic selection of features, such as forward selection and backward elimination. RapidMiner implements both of them with the operator “Optimize Selection”. Modify the previous model to include feature selection and choose the correct alternatives. Note that “Optimize Selection” is a nested operator. You must therefore define a Cross Validation operator within it with another classifier (e.g. decision tree) which will be used to select the features.



Evaluation Exercise IV

Data Management for
Digital Health, Winter
2019

Q25. Feature selection is an important step in the development of a clinical prediction model. [...]

After feature selection, the overall accuracy of the new model is similar to that of the previous one with the advantage of being substantially simpler

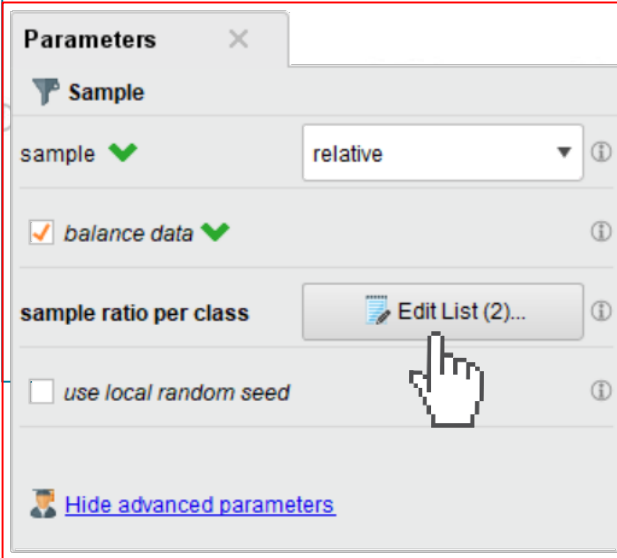
The root node of the new decision tree is the renal SOFA score

Creatinine values, potassium values, and liver disease are some of the key attributes

Hemoglobin levels are excluded from the new model

Q20. Feature selection is an important step in the development of a clinical prediction model [..]

Model with "Optimize Selection"




Parameters ✕


Sample

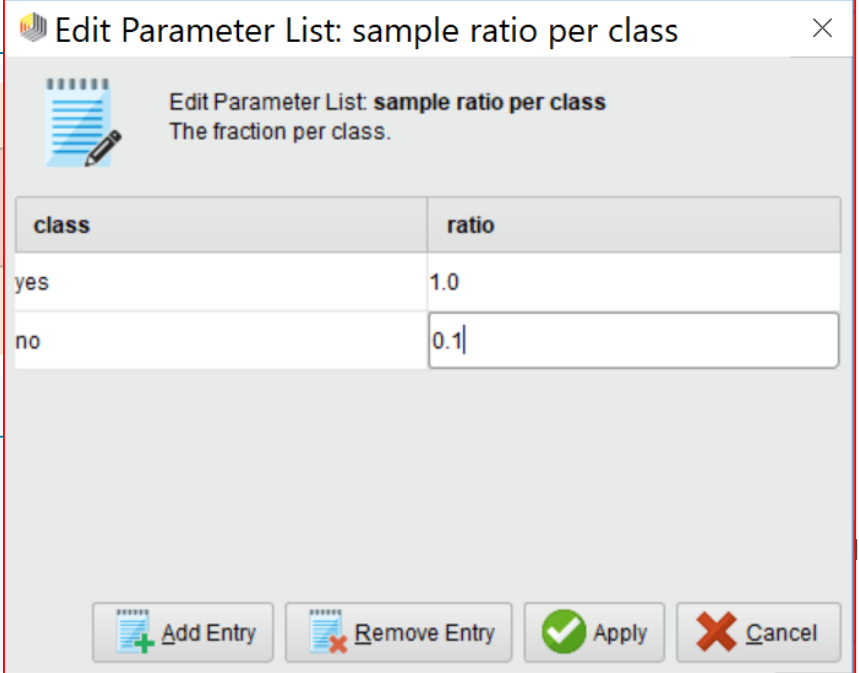
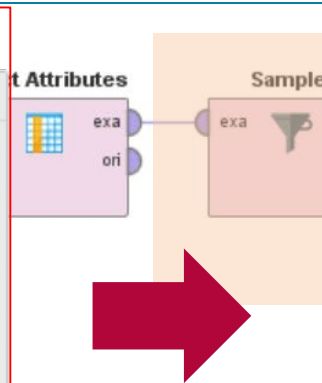
sample ✓ relative ⓘ

balance data ✓ ⓘ


sample ratio per class  Edit List (2)... ⓘ

use local random seed ⓘ





 [Hide advanced parameters](#)



Edit Parameter List: sample ratio per class ✕

 Edit Parameter List: **sample ratio per class**
The fraction per class.

class	ratio
yes	1.0
no	0.1

 Add Entry  Remove Entry  Apply  Cancel

Q20. Feature selection is an important step in the development of a clinical prediction model [..]

Model with "Optimize Selection"

attribute	wei... ↓
ELIXHAUSER_VANWALRAVEN	1
RHEUMATOID_ARTHRITIS	1
SOFA_RENAL	1
ALBUMIN_MAX	1
GLUCOSE_MIN	1
PTT_MAX	1

```
SOFA_RENAL ≤ 0.500
| ELIXHAUSER_VANWALRAVEN = ?: no {no=8, yes=0}
| ELIXHAUSER_VANWALRAVEN > 20.500
| | RHEUMATOID_ARTHRITIS > 0.500: yes {no=0, yes=14}
| | RHEUMATOID_ARTHRITIS ≤ 0.500
| | | PTT_MAX > 29.350
| | | | ALBUMIN_MAX = ?
| | | | PTT_MAX > 32.700
| | | | | GLUCOSE_MIN > 106.500: yes {no=0, yes=9}
| | | | | GLUCOSE_MIN ≤ 106.500
| | | | | | GLUCOSE_MIN > 99.500: no {no=2, yes=0}
| | | | | | GLUCOSE_MIN ≤ 99.500: yes {no=7, yes=21}
| | | | | PTT_MAX ≤ 32.700: no {no=2, yes=0}
| | | | | ALBUMIN_MAX > 3.200: no {no=3, yes=0}
| | | | | ALBUMIN_MAX ≤ 3.200
| | | | | PTT_MAX > 49.450: no {no=3, yes=0}
| | | | | PTT_MAX ≤ 49.450: yes {no=1, yes=8}
| | | | PTT_MAX ≤ 29.350: yes {no=0, yes=6}
| ELIXHAUSER_VANWALRAVEN ≤ 20.500
```


Q25. Feature selection is an important step in the development of a clinical prediction model. [...]

After feature selection, the overall accuracy of the new model is similar to that of the previous one with the advantage of being substantially simpler



The root node of the new decision tree is the renal SOFA score



Creatinine values, potassium values, and liver disease are some of the key attributes



Hemoglobin levels are excluded from the new model

