

Exercise 3 (18 points) – individual work

- The answers can be typed or handwritten (handwriting must be clear and readable), in this exercise sheet or your own sheet (put your name & ID at the top of the sheet). All answers must be saved to only 1 PDF file.
- Some questions also require the submission of processes/workflows (file.rmp or file.ipynb).
- In case of re-submission (after first grading) or submission after solution is given, your points will be weighted by 0.5.

1. (Total 5 points) Use training data in the given table for Naïve Bayes classification.

Predict a target record (F, F, positive):

Record	A	B	C	Buy (class)
1	T	T	positive	yes
2	F	T	negative	no
3	T	F	neutral	yes
4	T	F	negative	no
5	F	T	neutral	no
6	T	T	positive	yes
7	F	T	negative	no
8	T	F	positive	yes
9	F	F	negative	no
10	T	T	neutral	yes

1.1 (2 points) Without Laplace correction

(a) Normalized probability to predict "yes" (answer in 2 decimal places) =

(b) Normalized probability to predict "no" (answer in 2 decimal places) =

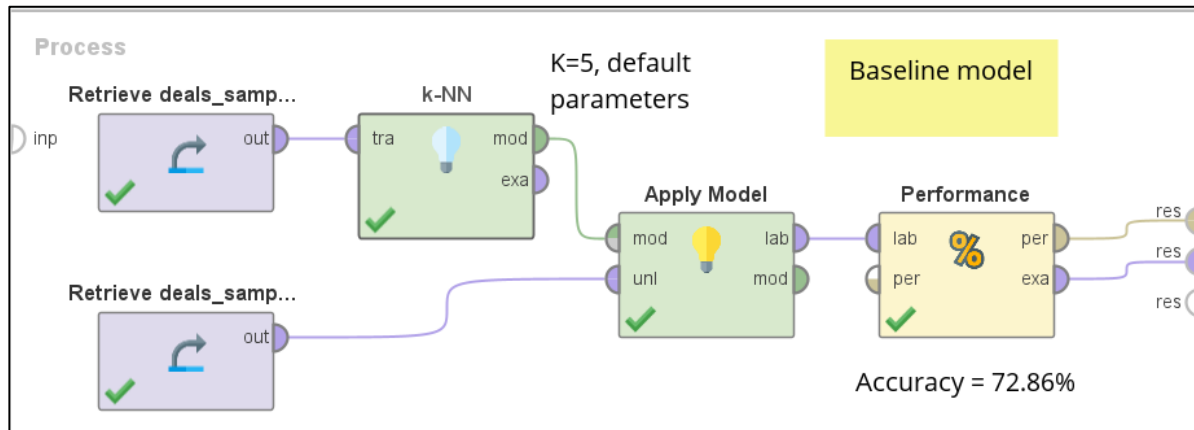
1.2 (3 points) With Laplace correction

(a) Normalized probability to predict "yes" (answer in 2 decimal places) =

(b) Normalized probability to predict "no" (answer in 2 decimal places) =

(c) Final prediction =

For questions 2 and 3, retrieve **deals-samp-train** for training and **deals-samp-test** for testing. Consider the base line model that uses KNN with default parameters ($K=5$) to predict future customer (yes or no). Its accuracy = 72.86%.



2. (Total 4 points) Built your workflow that also uses KNN to predict future customer. But add necessary preprocessing and adjust parameters to get better accuracy than the baseline.

2.1 (1 point) Submit your workflow. It must output at least performance table. Name the workflow **question2.rmp**.

2.2 (3 points) Answer whether each attribute requires any preprocessing. If it does, briefly explain which preprocessing is applied to it.

Attribute	Preprocessing? (YES + explain, or NO)
(a) Age	
(b) Gender	
(c) Payment Method	

3. (Total 9 points) Use neural network and SVM to classify future customer instead. Apply necessary preprocessing such that attributes are suitable for both classifiers (assuming the same preprocessing pipeline for both classifiers). Adjust parameters to get better accuracy than the baseline.

3.1 (2 points) Submit your workflow(s) for both classifiers. The workflow for each classifier must output at least performance table and classification model. Name the workflow **question3.rmp**.

3.2 (3 points) Answer whether each attribute requires any preprocessing. If it does, briefly explain which preprocessing is applied to it.

Attribute	Preprocessing? (YES + explain, or NO)
(a) Age	
(b) Gender	
(c) Payment Method	

3.3 (2 points) Check your neural network model. Report number of input nodes that represent each attribute, number of hidden layers, and size of each hidden layer.

Note 1: Your neural network should have at least 1 hidden layer.

Note 2: Try a few different numbers of hidden layers, with a few different #nodes per layer. Take note of their performance & choose the best option.

(a) #input nodes for Age =
(b) #input nodes for Gender =
(c) #input nodes for Payment Method =
(d) #nodes in hidden layer 1 =
<i>If using >1 hidden layer, also report #nodes in hidden layers 2, 3, ...</i>

3.4 (2 points) Check your SVM kernel model.

Note: Try simple kernels such as dot, radial, or polynomial. Other complicated kernels may take a long time to run.

(a) List attributes with positive weights =
(b) List attributes with negative weights =