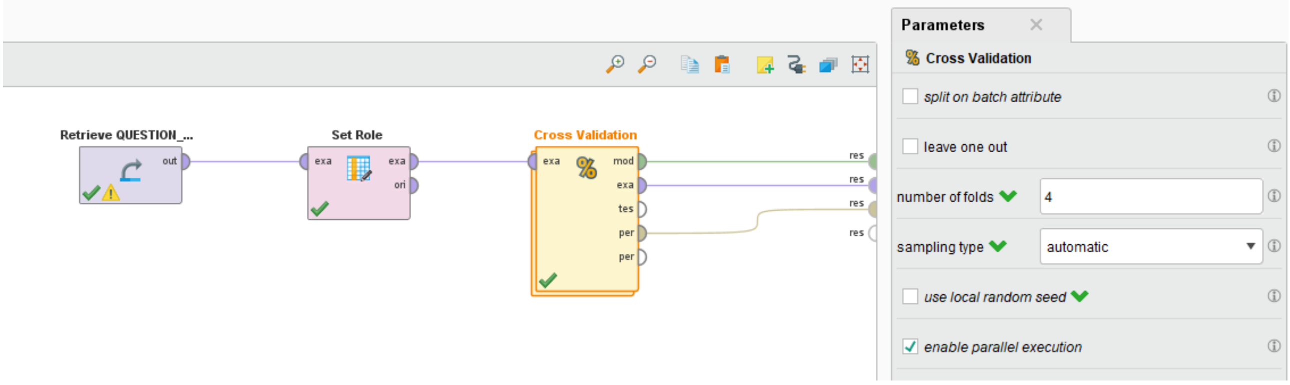
Question 1

Question 2

(a)

After importing the situation table from Celonis into RapidMiner we can design the decision tree learning process using the cross validation operator after setting the class variable:

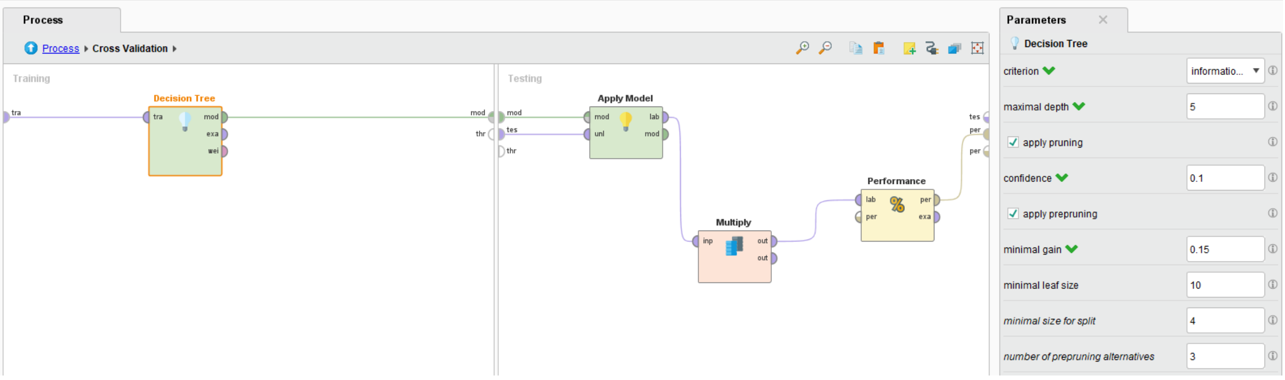


Parameters

Cross Validation

- □ split on batch attribute
- □ leave one out
- number of folds: 4
- sampling type: automatic
- □ use local random seed
- ☑ enable parallel execution

We construct the training section of the cross validation using the decision tree operator in RapidMiner and also specifying the given parameters. In the testing section we simply apply the current test set to the given model and run a performance check upon it.

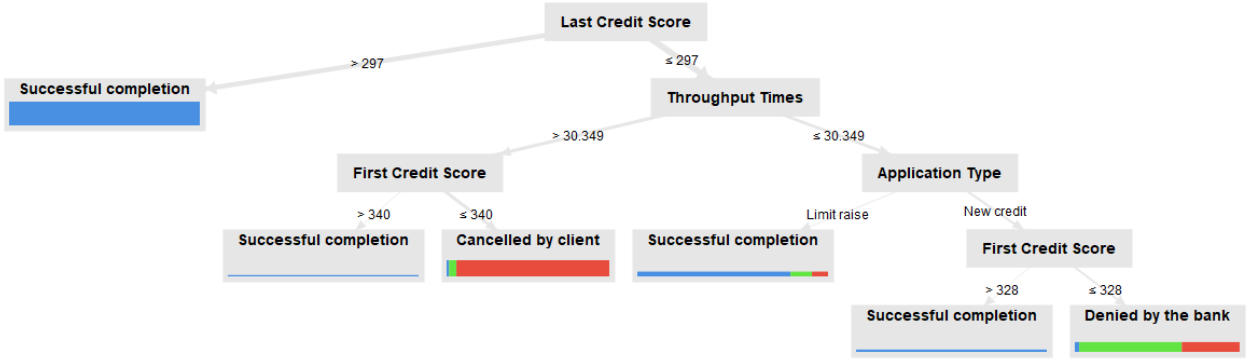


Parameters

Decision Tree

- criterion: informatio...
- maximal depth: 5
- ☑ apply pruning
- confidence: 0.1
- ☑ apply prepruning
- minimal gain: 0.15
- minimal leaf size: 10
- minimal size for split: 4
- number of prepruning alternatives: 3

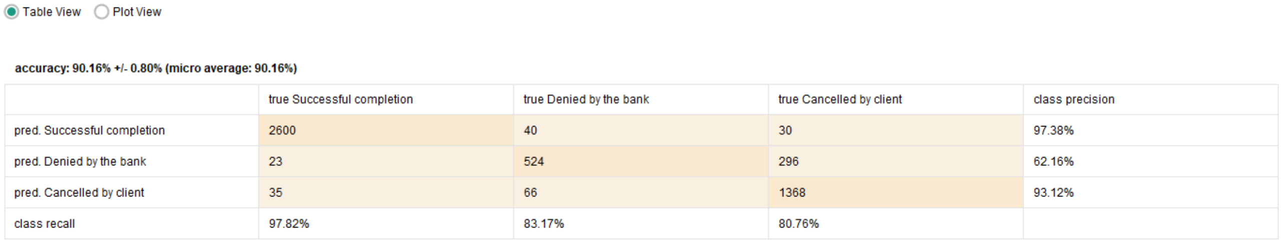
This process results in the following decision tree:



Here we can see that for all applications, where the applicants last credit score was above 297, the application was successfully completed.

We can also observe, that the very few applicants for new credits, those credit scores dropped significantly (First Credit Score > 329 and Last Credit Score ≤ 297) within a time less or equal to 30.349 days, all completed their application successfully.

Using the cross validation operator, we could also observe the following performance metrics:



● Table View ○ Plot View

accuracy: 90.16% +/- 0.80% (micro average: 90.16%)

	true Successful completion	true Denied by the bank	true Cancelled by client	class precision
pred. Successful completion	2600	40	30	97.38%
pred. Denied by the bank	23	524	296	62.16%
pred. Cancelled by client	35	66	1368	93.12%
class recall	97.82%	83.17%	80.76%	

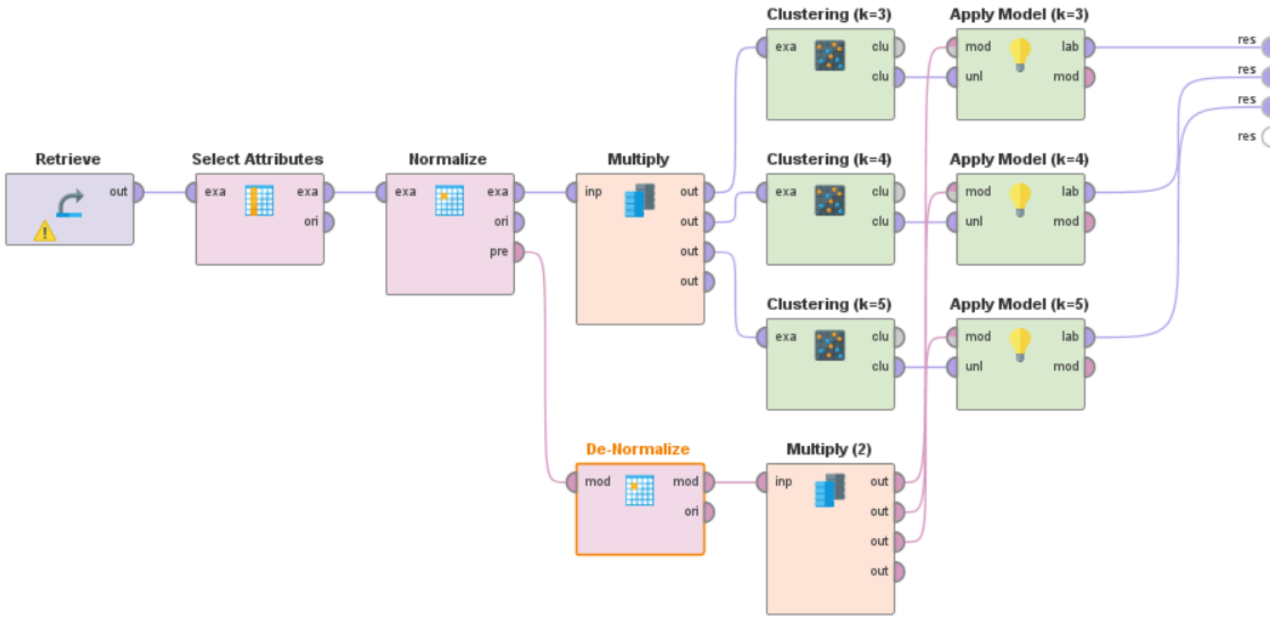
(b)

1. Most applicants with a last credit score of at most 297 cancelled their own application when it was already roughly over a month's time in progress. Perhaps the current application process needs to be revised to minimize throughput times.
2. 62% of applications, that was roughly under a months time in progress, by applicants with a last credit score of at most 297 for a new credit were rejected by the bank, if their first credit score was at most 328. If the bank seeks for more successful application completions, perhaps they should adjust their rejection criteria.

Question 3

(a)

After selecting the three specified attributes, normalization by Z-transformation is done. The result is multiplied so it can be fed to all k-means clusterers at once. For each clustering we set the necessary 'k' and 'max runs' and select the 'add as label' checkbox so the cluster labels are added as a new column. To use the unscaled dataset we first reverse the Z-transformation by De-Normalizing and then apply it to the clustered data.



(b)

Question 4

(a)

(b)

Question 5

(a)

(b)

(c)

(d)