UNDERSTANDIN G THE CASE STUDY

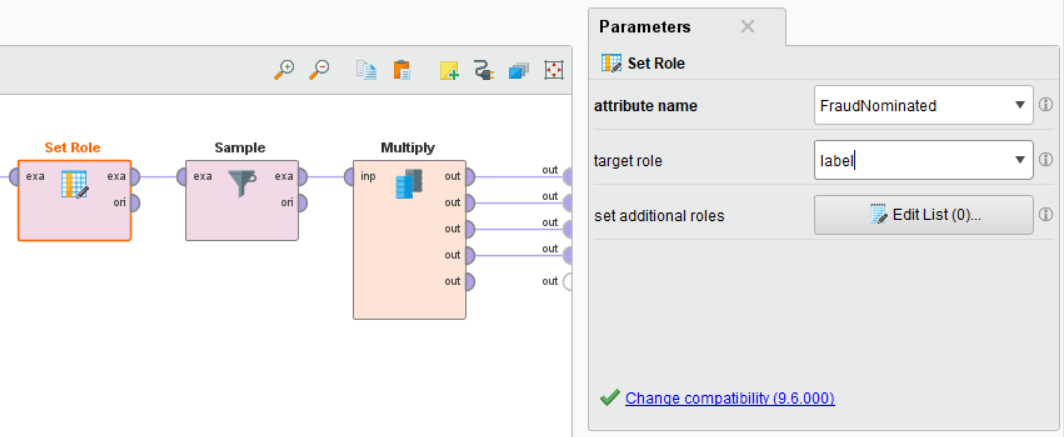
**Case Study Analysis:**

Test bank want to detect and investigate the fraud suspect for bank. The provide us historical data contain set of attributes include fraud and not fraud attribute. Our mission is to provide a satisfying report. It suggests that our developed model is able to detect 90% actual fraud with 70% accuracy on given dataset.

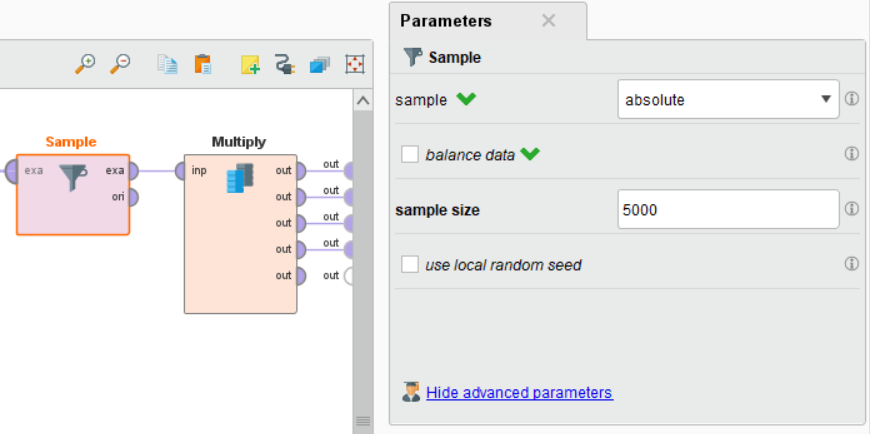
PROCESSING APPLIED

**Setting the role and rescaling:**

Set Role:



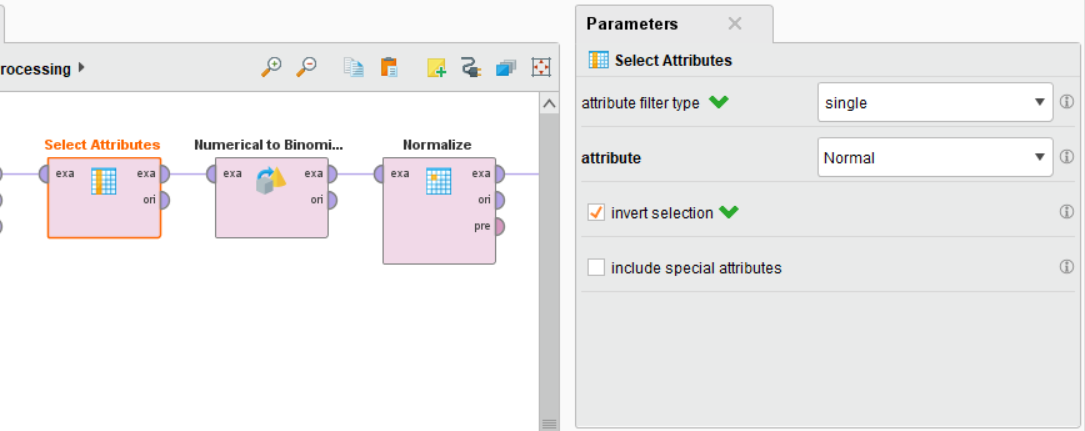
Rescaling:



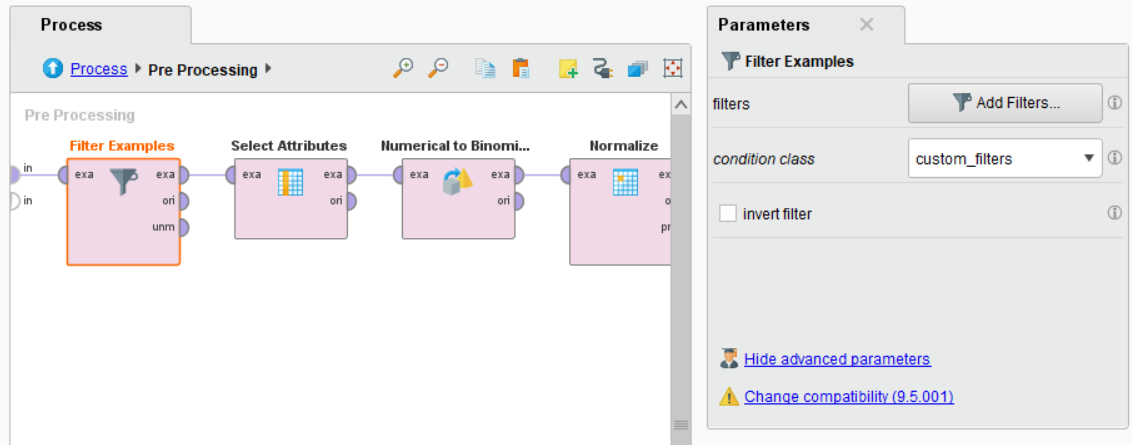
*(RESCALE: BECAUSE OF LAPTOP HAVE LOW CONFIGURATION.)*

**Removing synonymous and noisy attributes:**

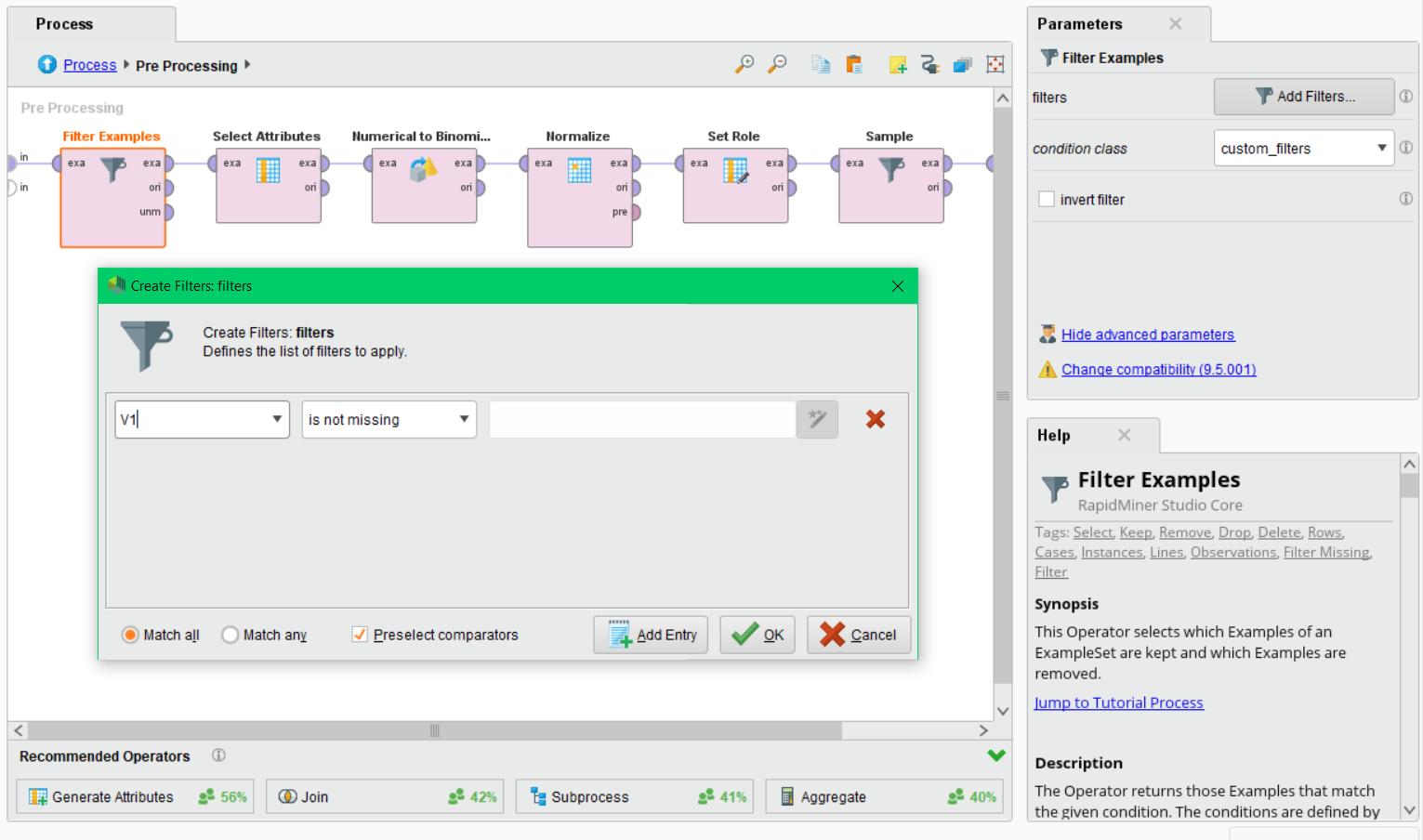
Removing synonymous:



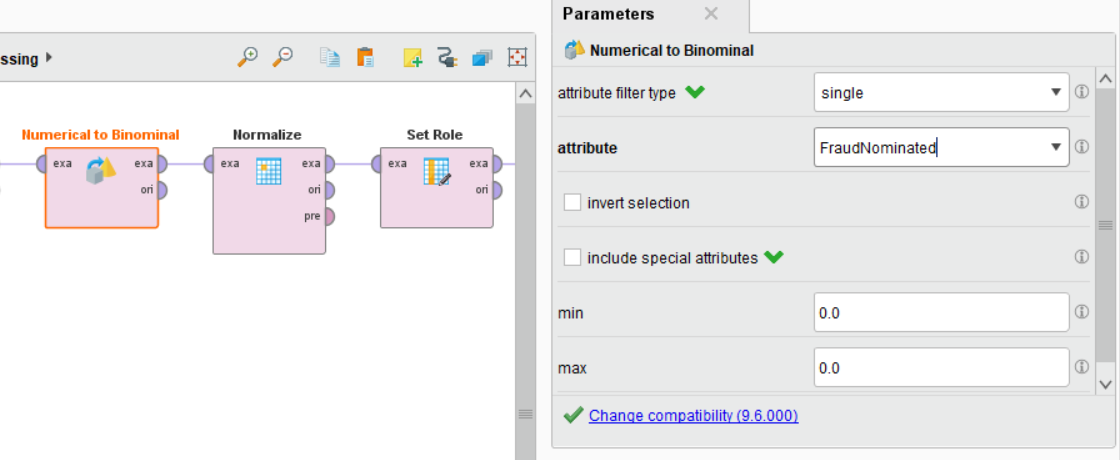
Removing noisy attributes:



**Dealing with missing values:**



**Negating the label for relevant ROC and Recall:**

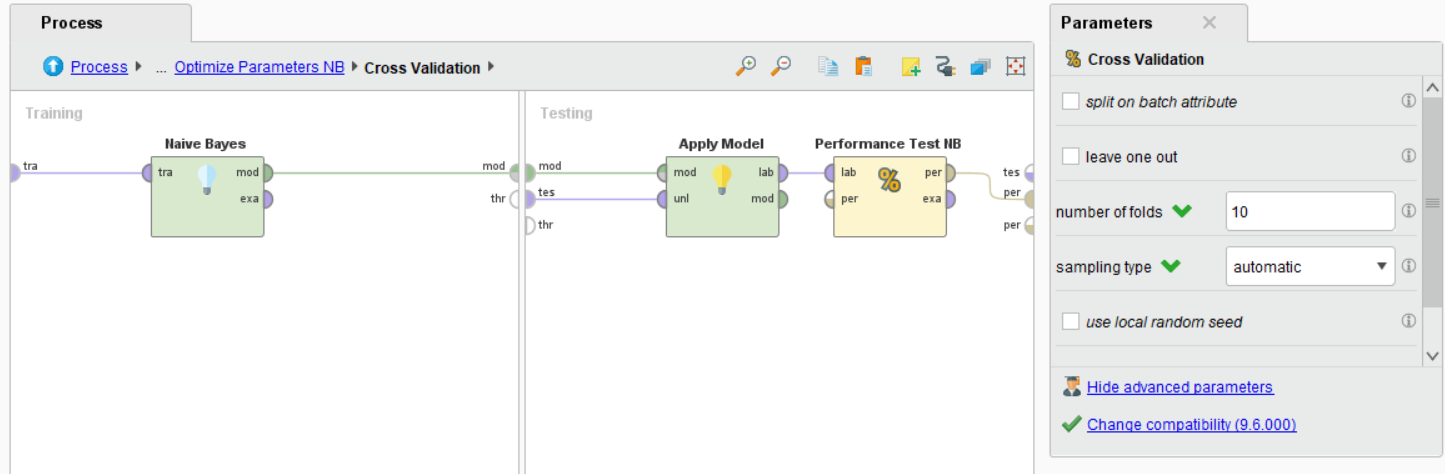


(It means, we have to change value Fraud nominated o,1 to yes, no. so to handle that Numerical to binomial operator allow us to convert)

APPLIYING FIRST MODEL OR TECHNIQUE

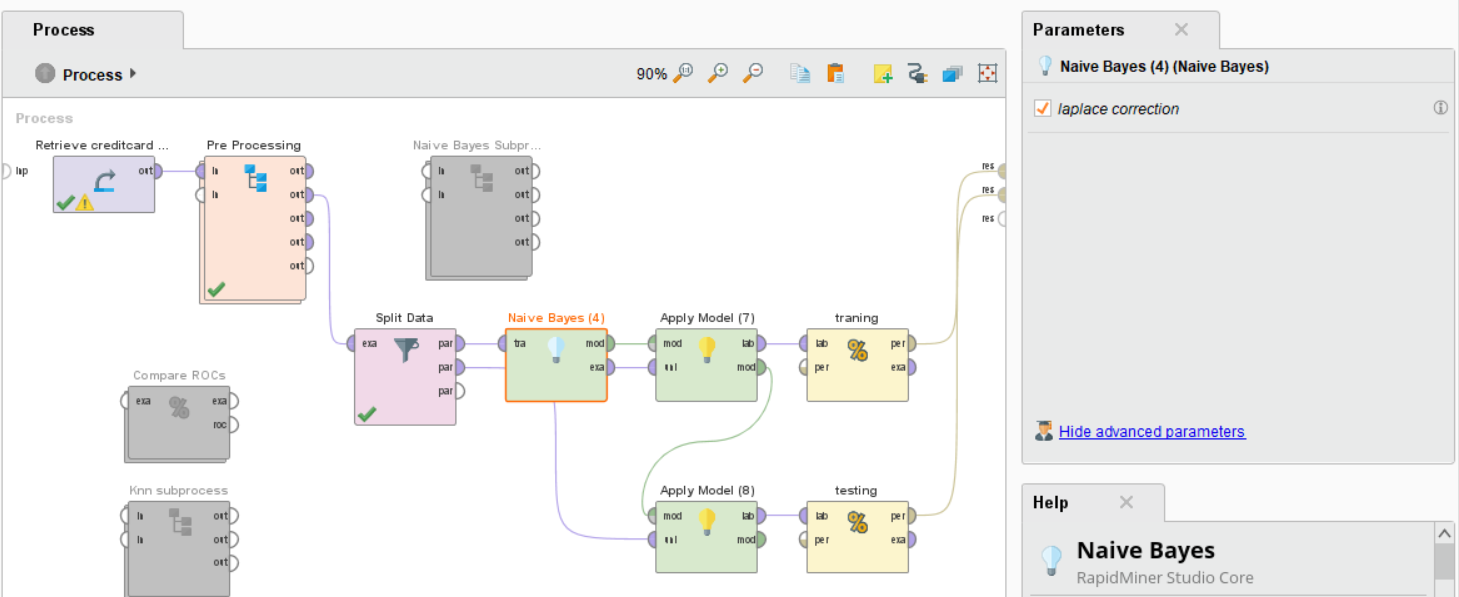
**The motivation of choosing Naïve Bayes is following:**

1. Handle continuous and discrete data.
2. Can used for both binary and multi-class classification problem.
3. Can make probabilistic prediction.
4. Need less training data and easy to implement

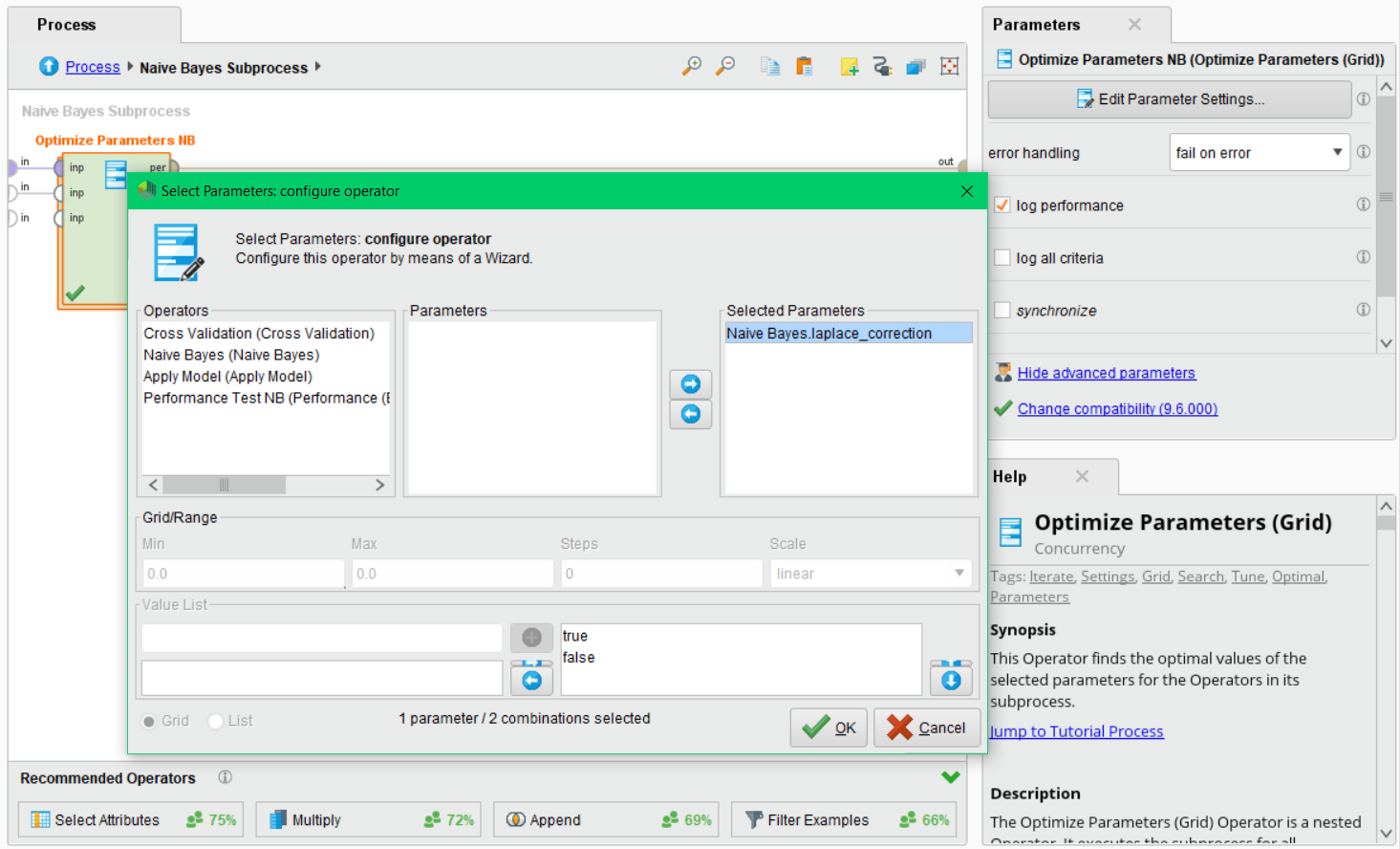


**Setting main parameter as Model Parameter:**

Default parameter setting is checked on Laplace correction. when uncheck parameter then accuracy is 97.60% and recall is 82.61%. When checked the setting its testing accuracy is increase at 97.66% and recall is 85.65%. when we add optimization technique it automatically set best value for optimization.

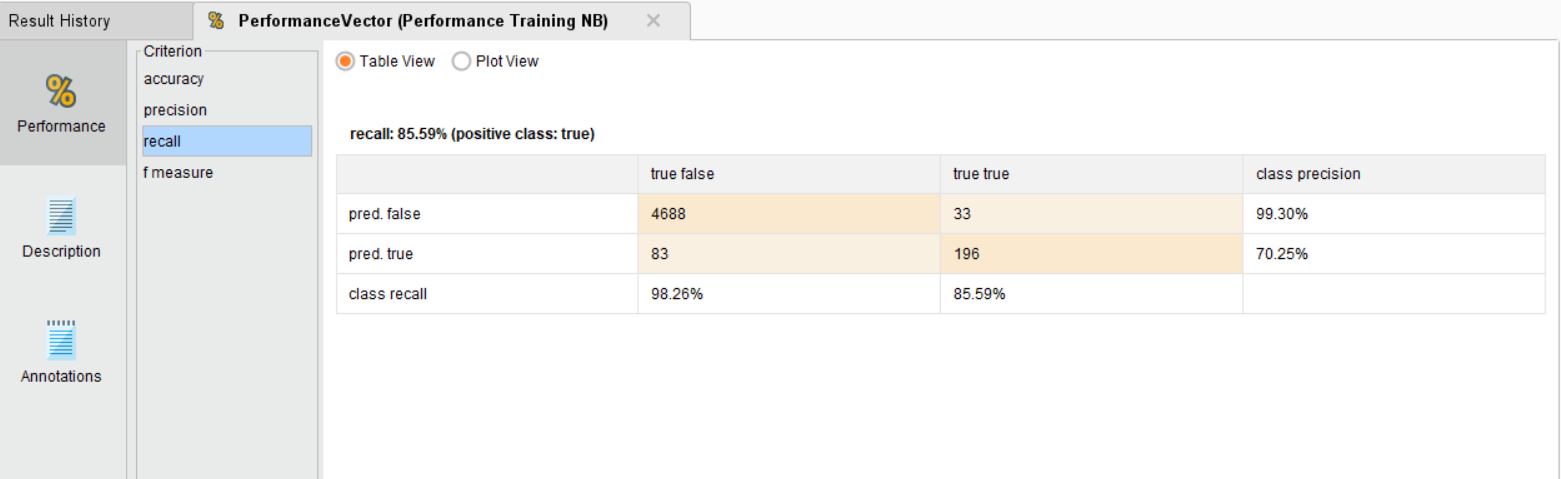


**Setting other parameters as optimization parameter:**

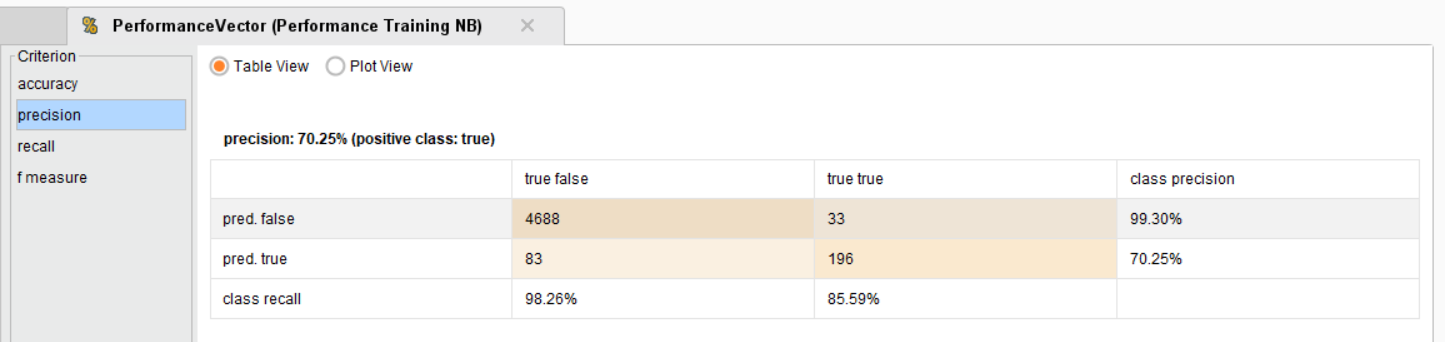


**Performance Results Recall and Precision of Training:**

Recall:



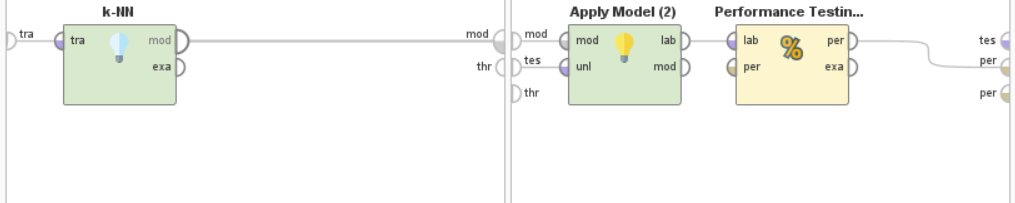
Precision:



APPLIYING SECOND MODEL OR TECHNIQUE

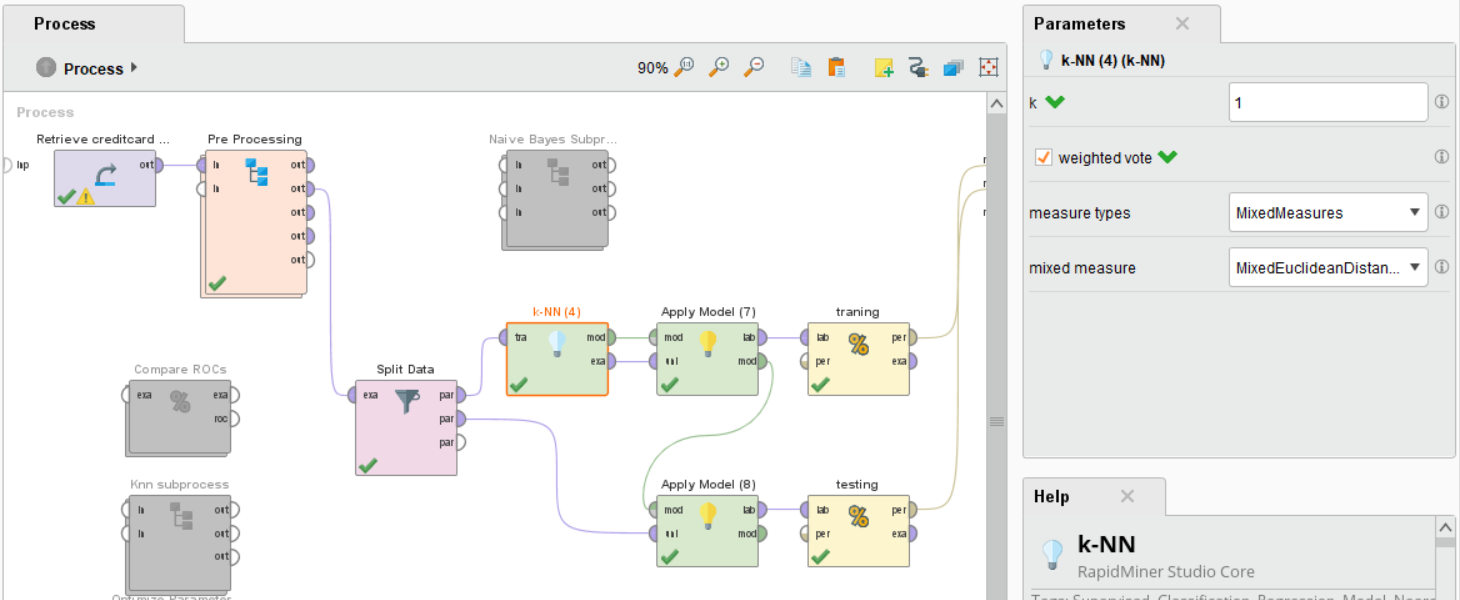
**The motivation of choosing k-nearest neighbors** **following:**

1. Effective if training data is large.
2. High accuracy, versatile and simple algorithm.
3. it is useful for classification and regression problem.
4. No assumption about data- no need to make additional assumption, just tune the several parameters and build fastest model.

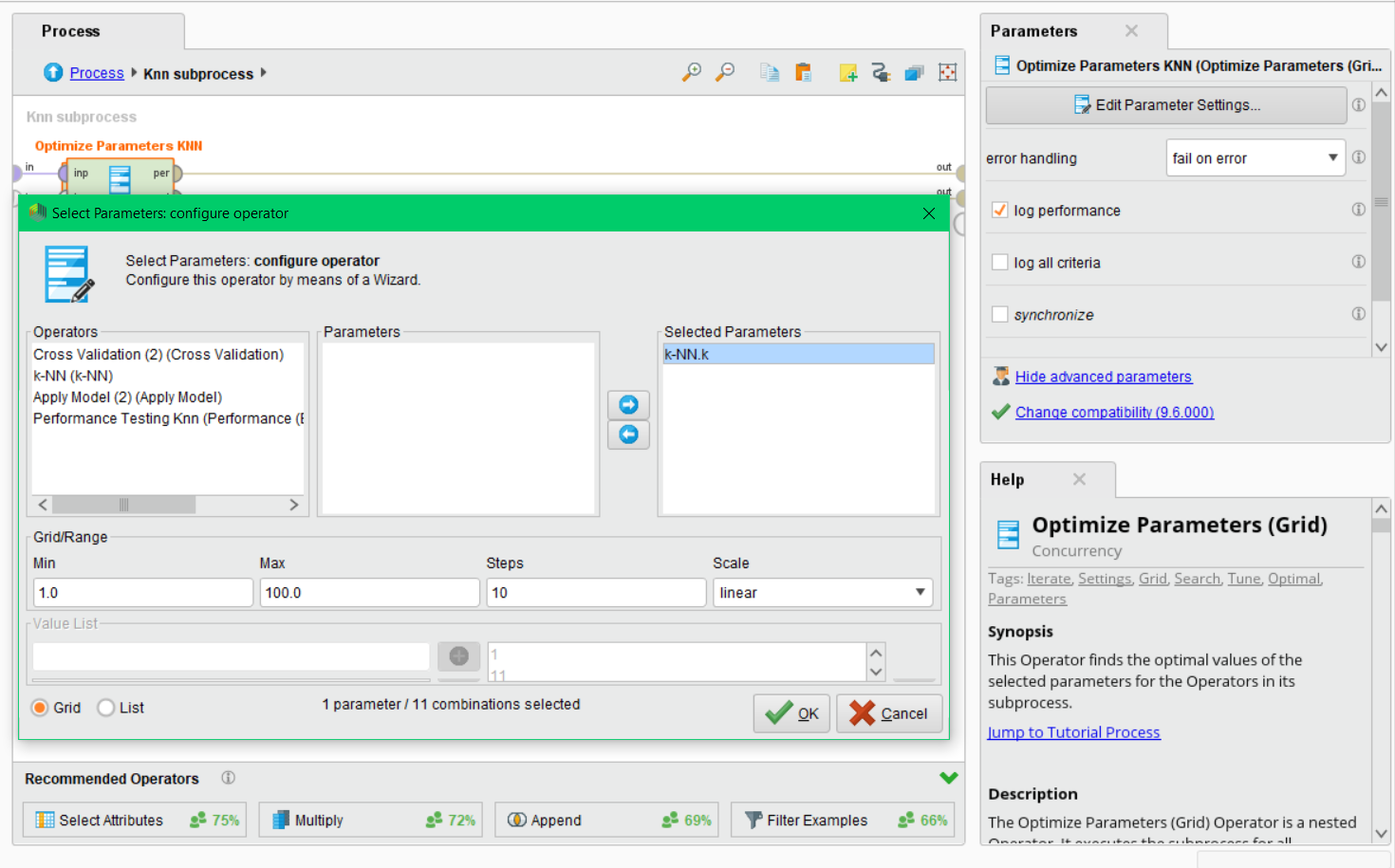


**Setting main parameters as model parameter:**

Default parameter setting of KNN is K=5 and weight voted is checked. when k=5 and weight vote is checked then accuracy is 99.00% and recall is 79.71%. When k=1 and weight vote is checked then its testing accuracy is increase at 99.07% and recall is 82.61%. when we add optimization technique it automatically set best value for optimization.

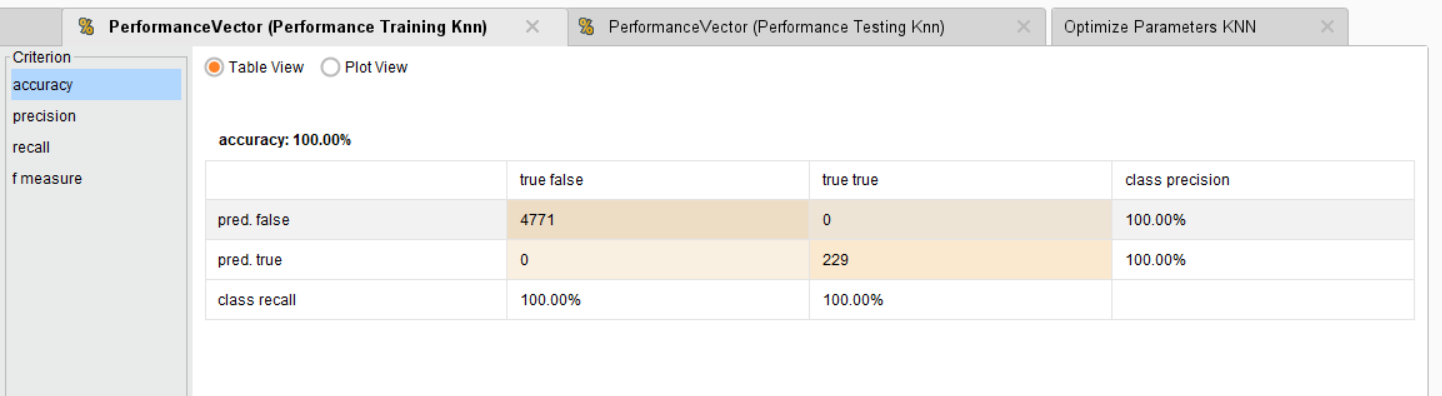


**Setting other parameters as optimization parameter:**

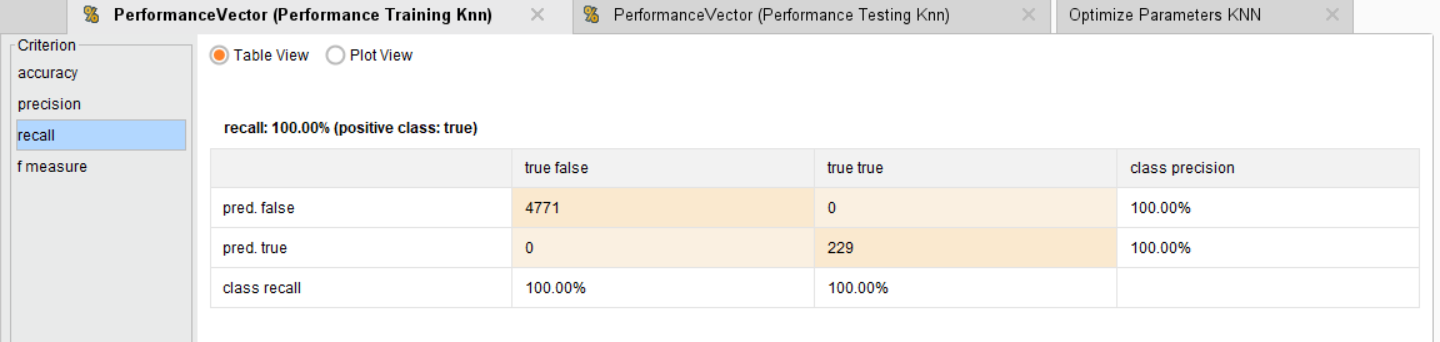


**Performance Results Recall and Precision of Training:**

Precision:

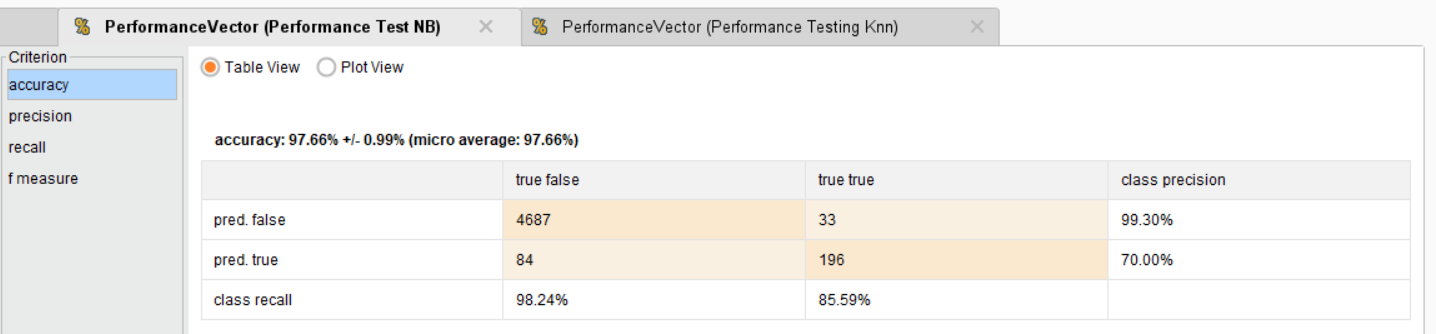


Recall:

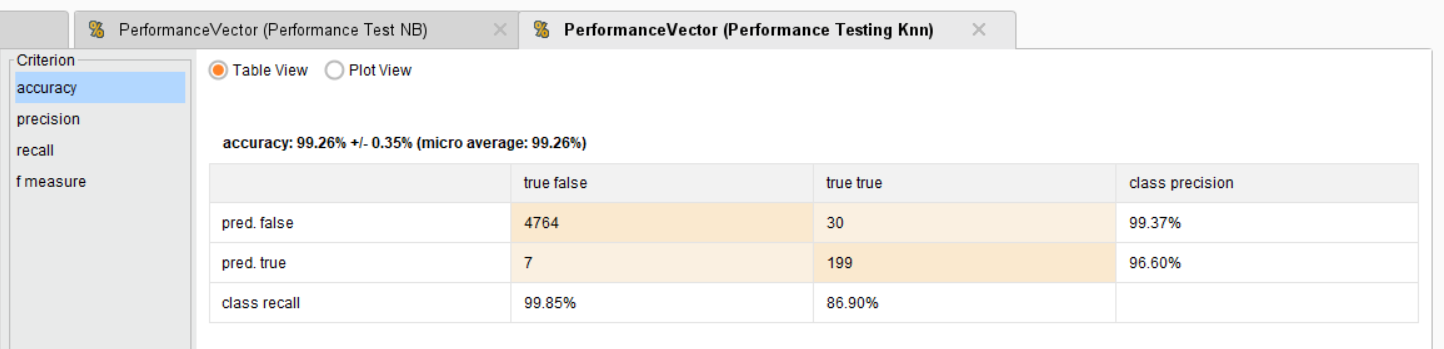


**COMPARISON OF PERFORMANCE S ON THE TEST DATASET:**

**Accuracy Testing:**

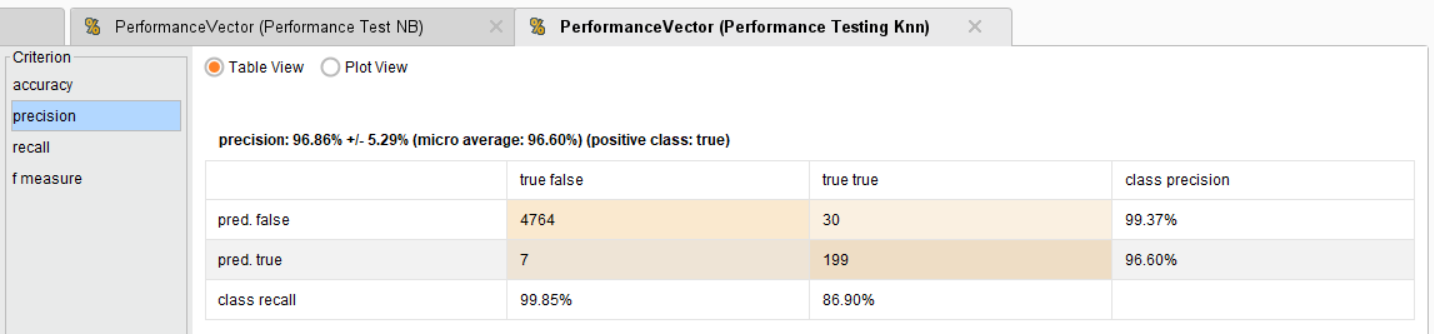


(Accuracy testing for naïve Bayes)

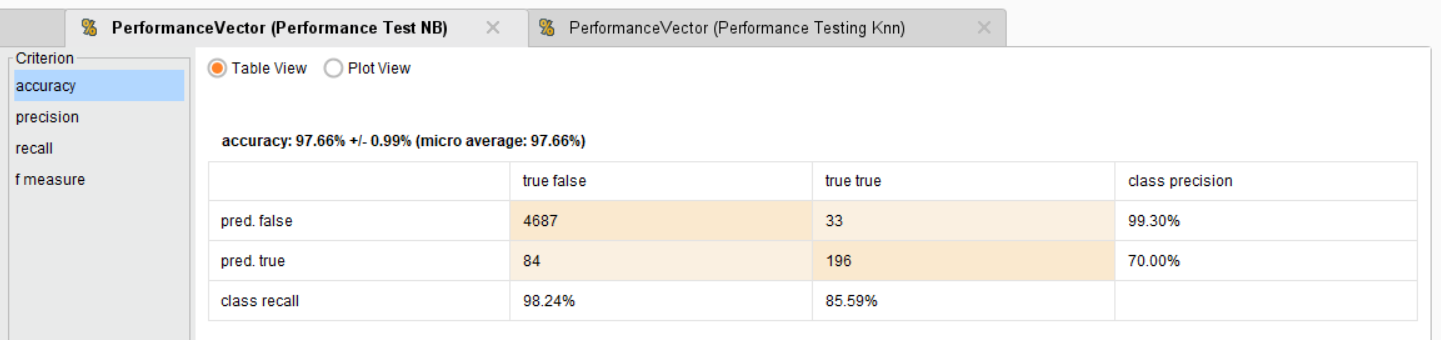


(Accuracy testing for KNN)

**Precision Testing:**

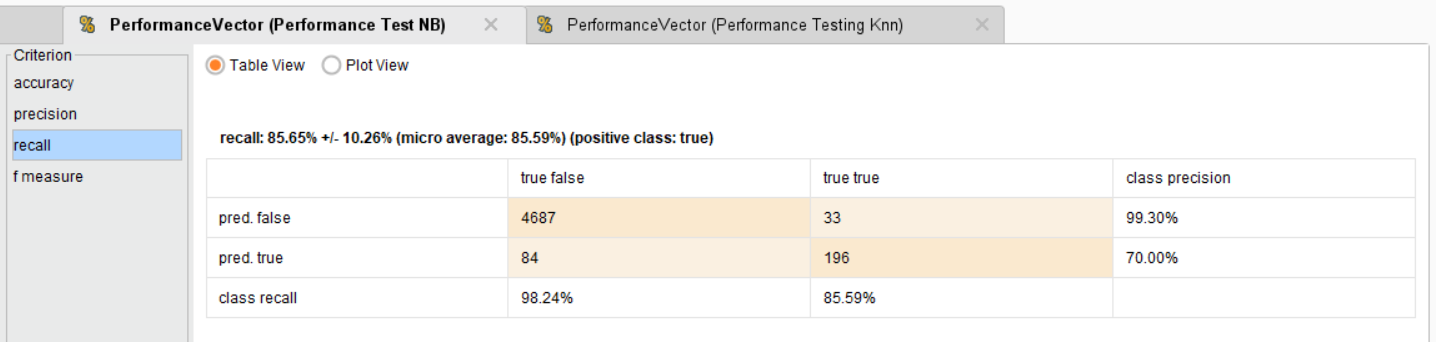


(precision testing for naïve KNN)

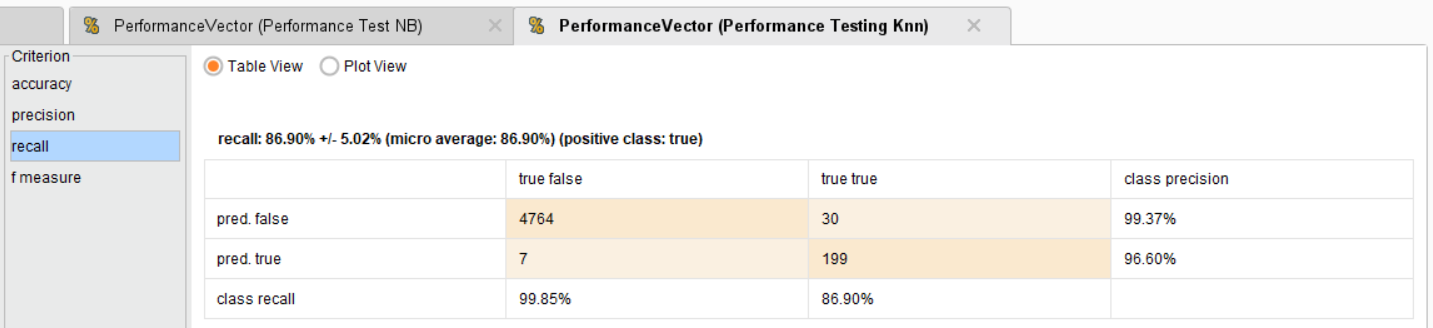


(precision testing for naïve Naive Bayes)

**Recall Testing:**

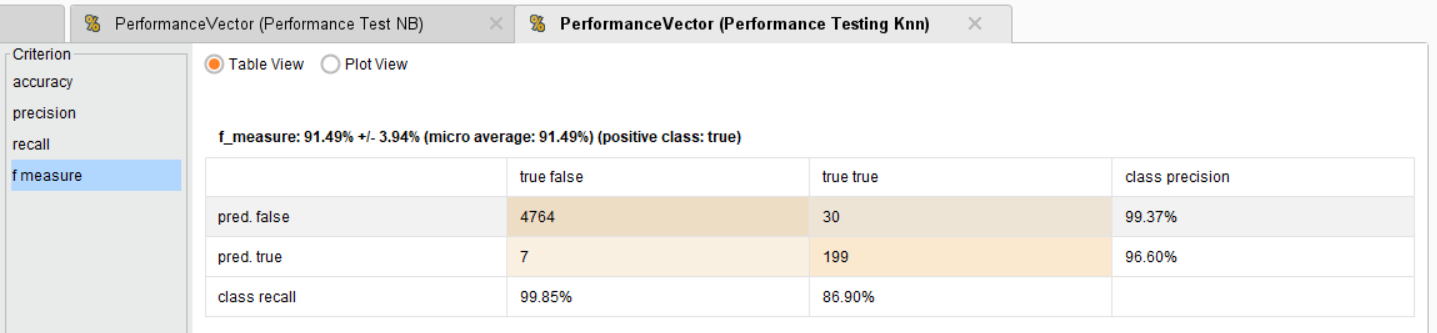


(Recall testing for naïve Naive Bayes)

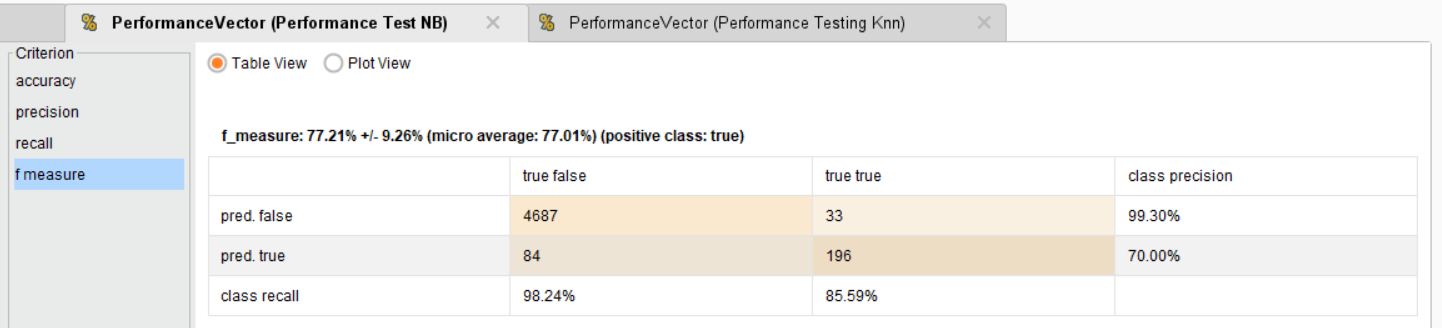


(Recall testing for KNN)

**F-measure Testing:**



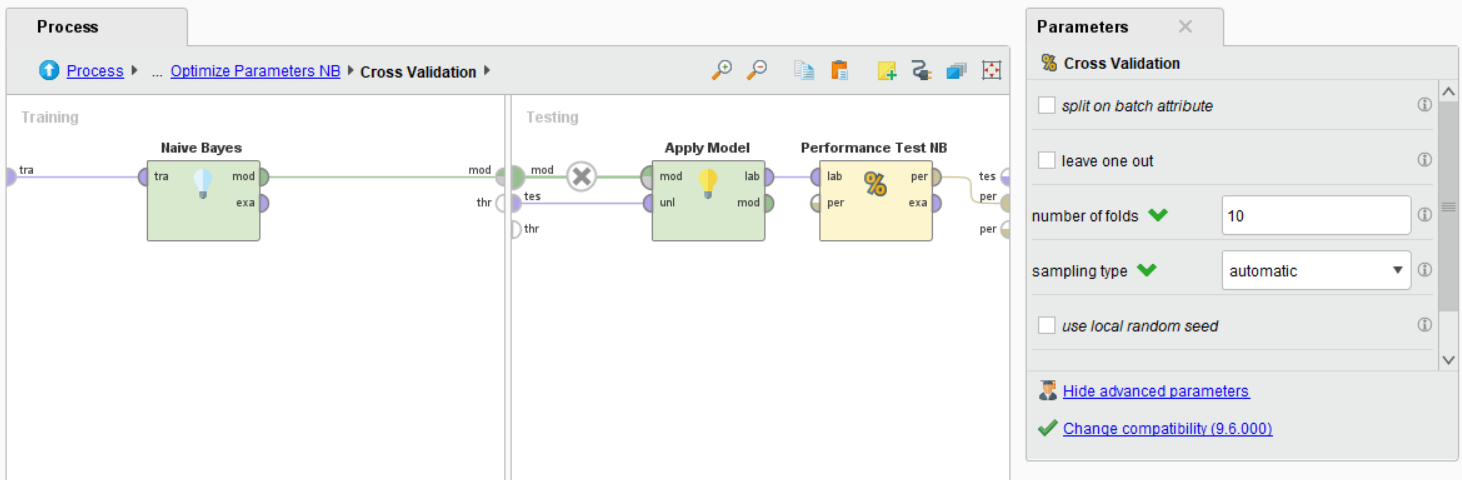
(F-Measure testing for KNN)



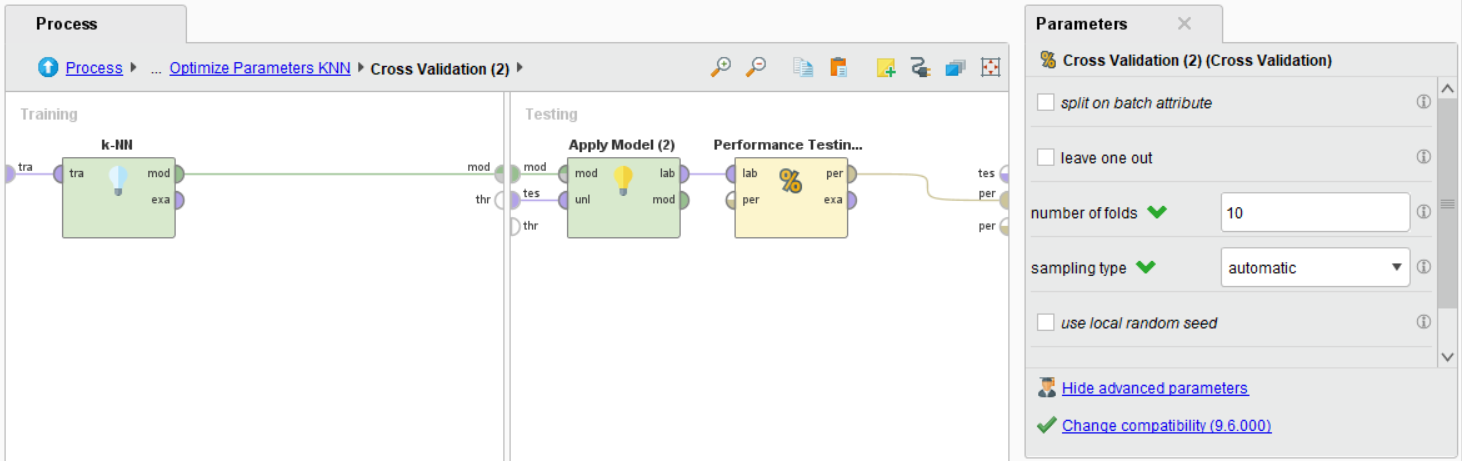
(F-Measure testing for Naïve Bayes)

**Use of Cross Validation for both techniques to deal with overfitting:**

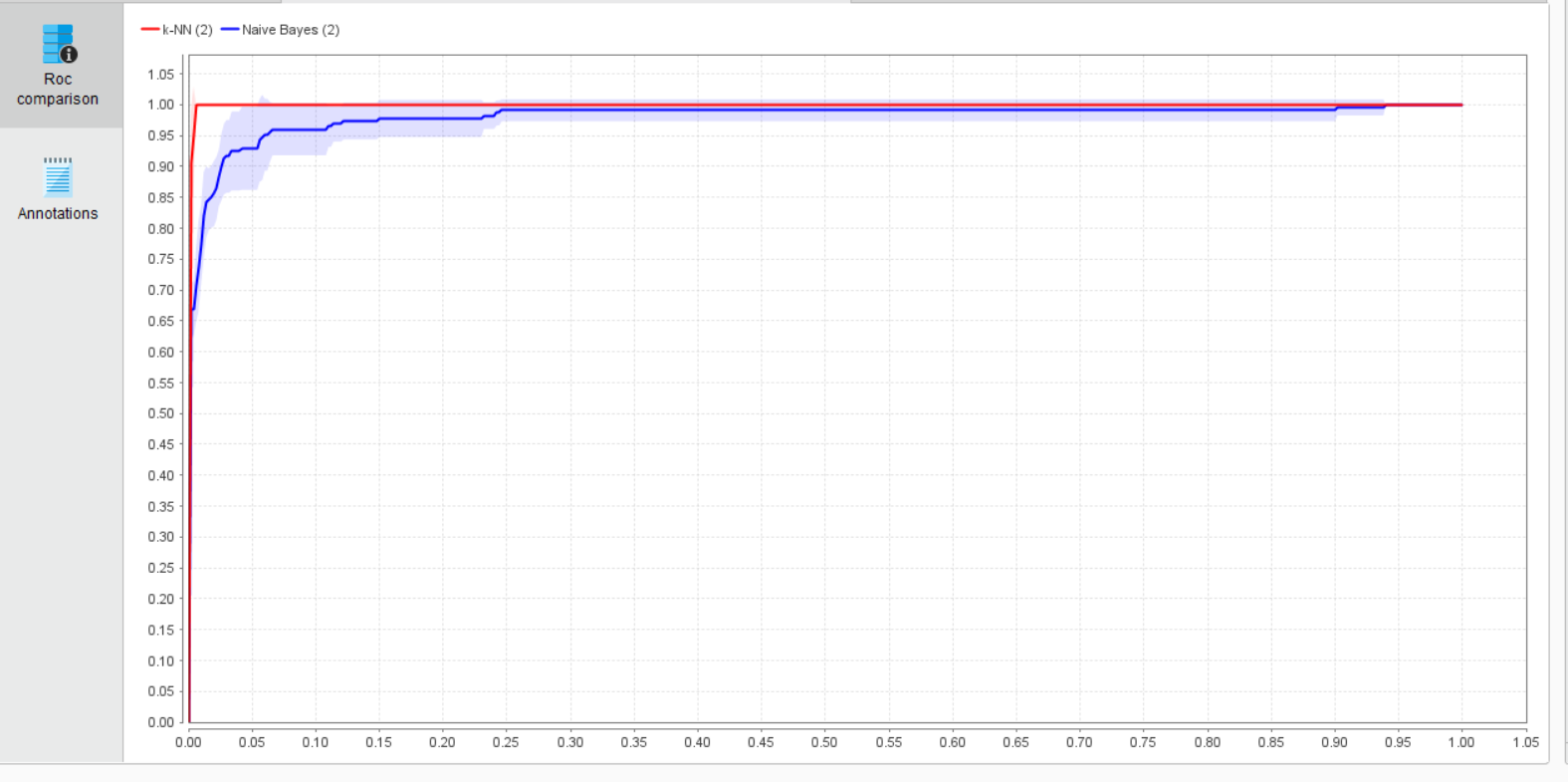
Naïve Byes Cross-Validation:



KNN Cross validation:



**Model selection (ROC):**



**FINAL RECOMMENDA TION OF BEST MODEL:**

**Technical perspective- overfitting discussion:**

On technical precative KNN is best fit model for this case because result of accuracy, precision, Recall, F-measure KNN has highest in all result comparison of naïve Bayes. In case of overfitting we use cross validation but outside of cross validation also it is not overfitting.

**Business perspective:**

Testing result have less false negative rate which decrease the possibility of loss money. So this is helpful for business perspective.

**FULL PROCESS:**

