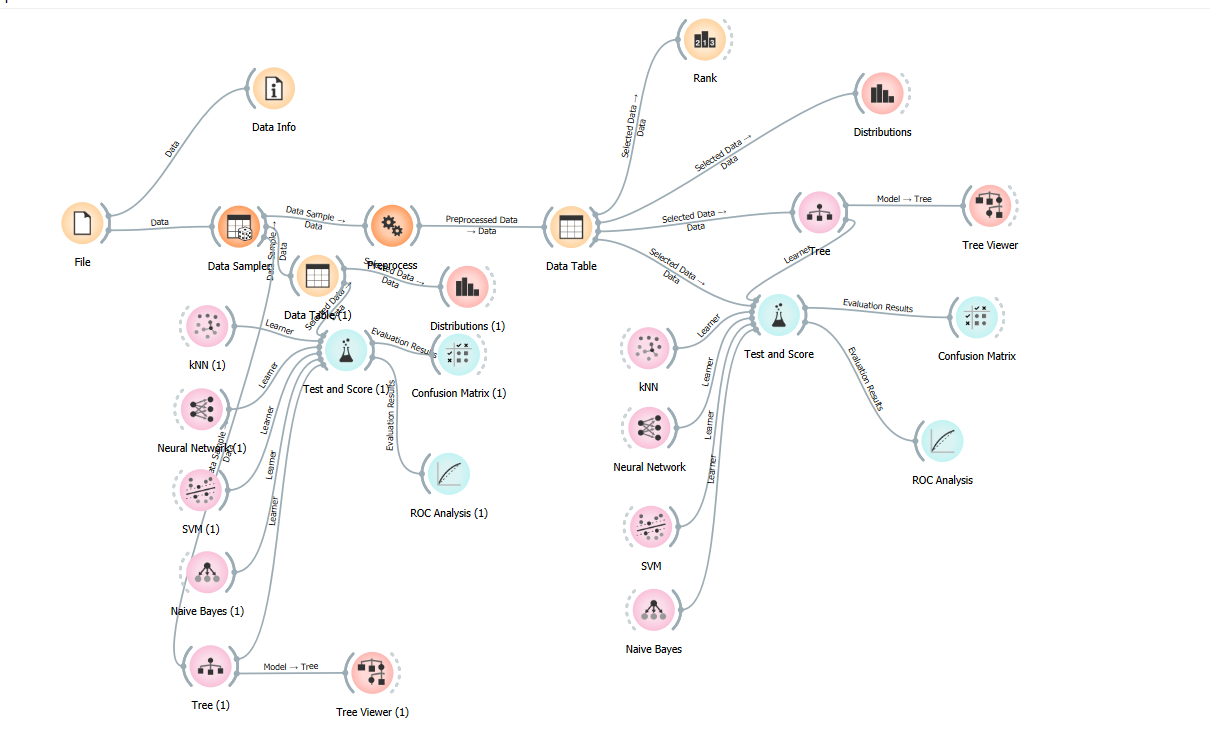
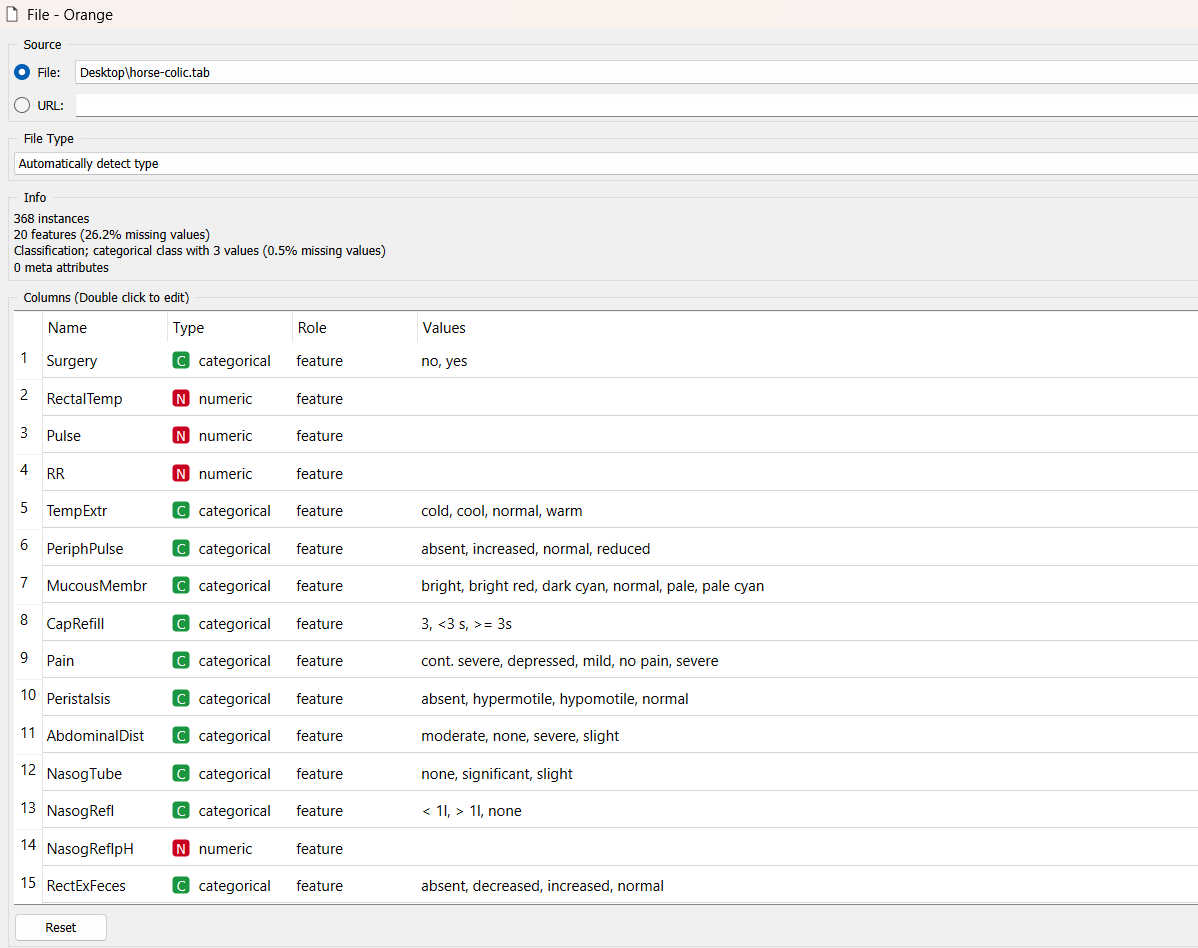
DATA MINING PROJECT

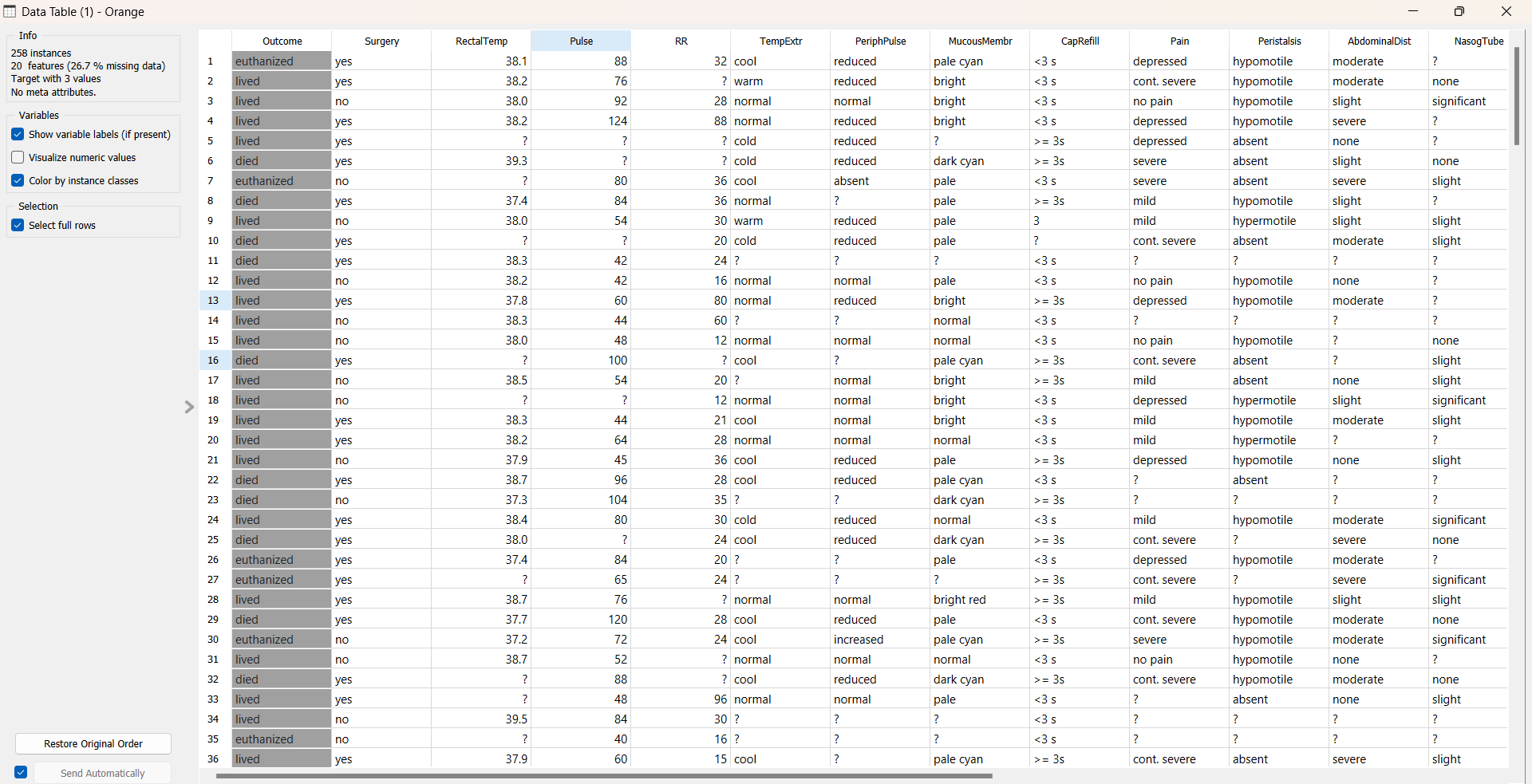
BATCH-18

horse-colic.tab

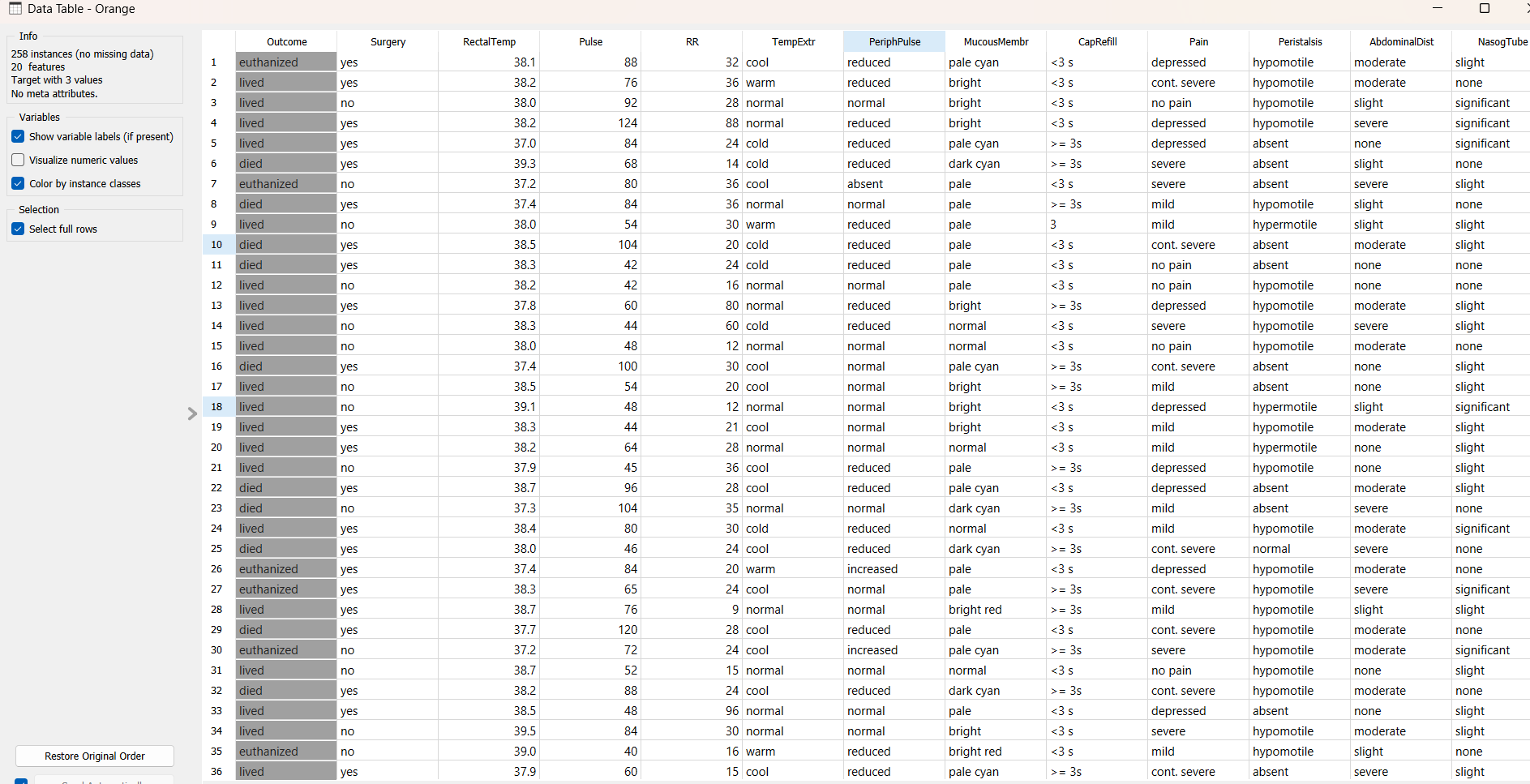
Dataset info:



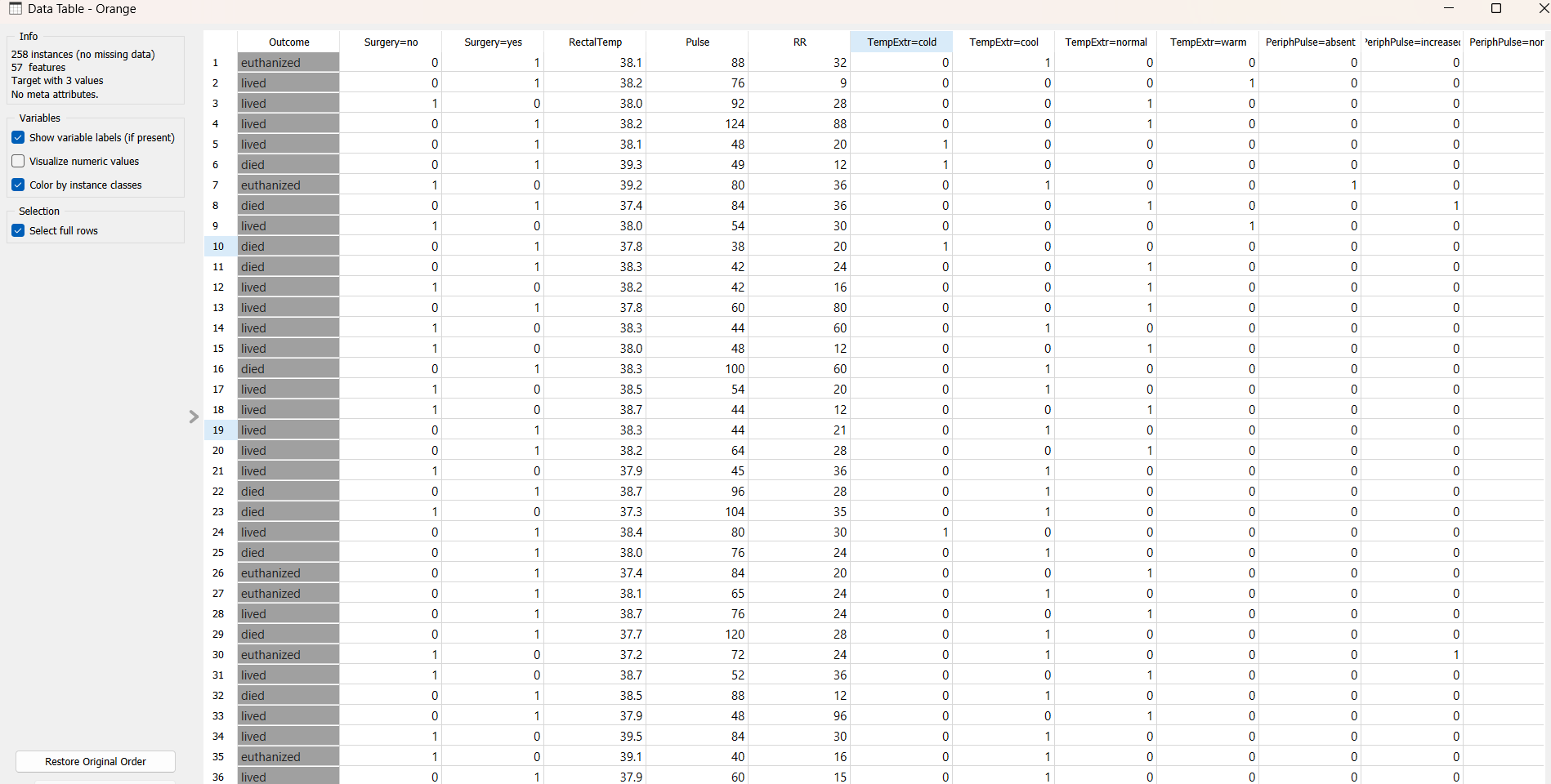
Step1 :Preprocessing



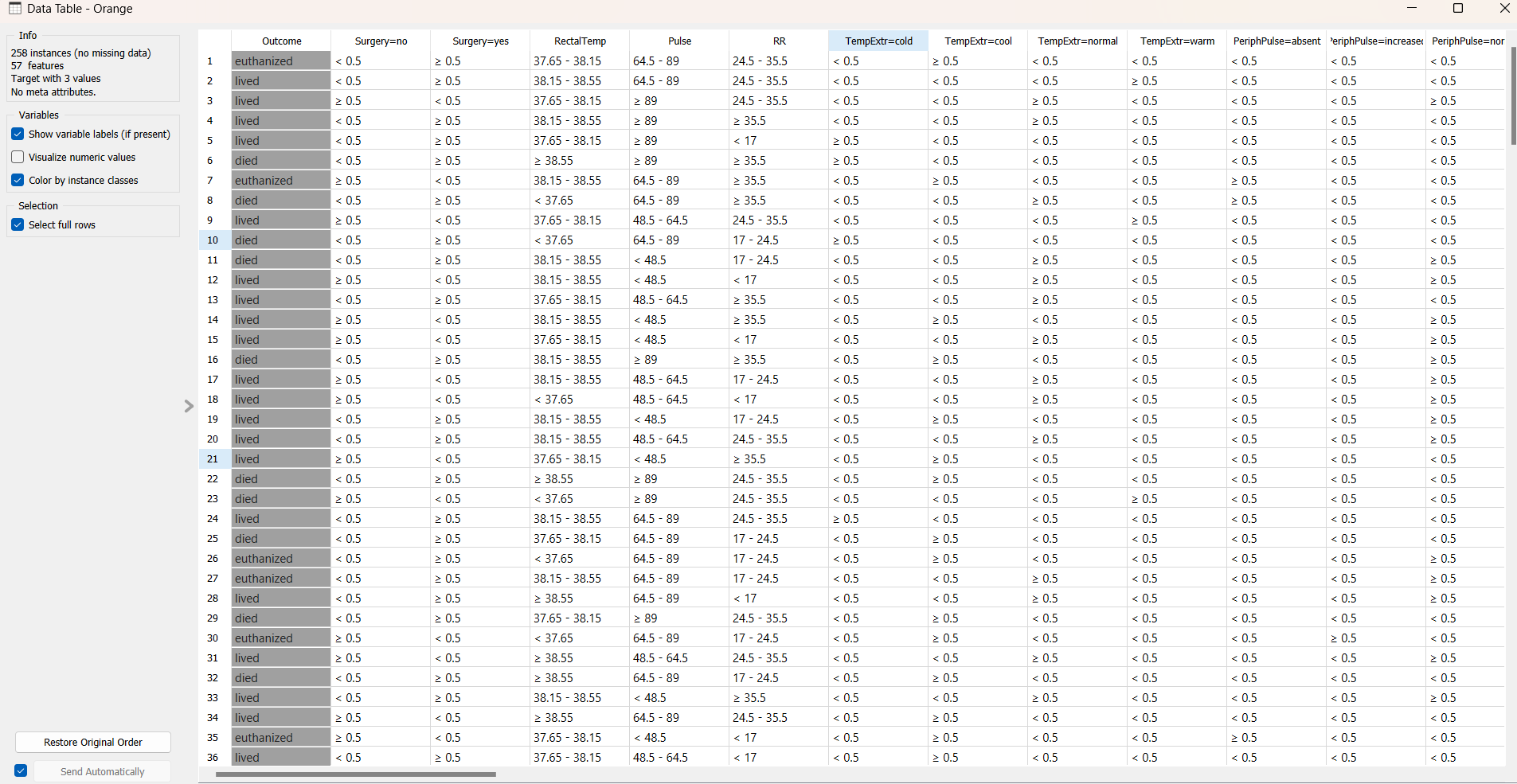
Impute missing values : Replace with random value



Continuous discrete variables : One feature per value



Discrete Continuous Variables: Equal frequency discretization



Step2 : We opted classification technique

1.KNN

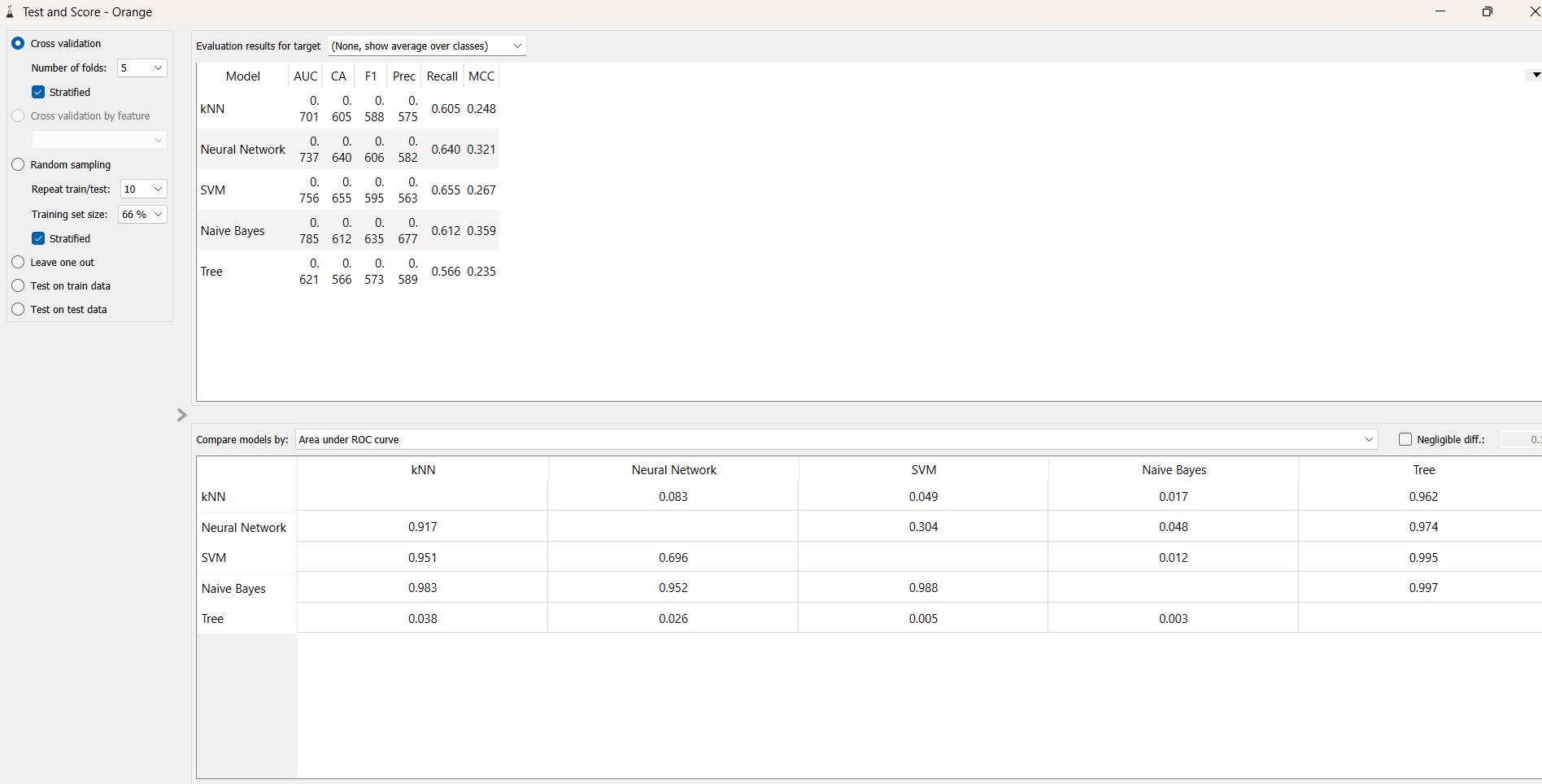
2.Neural Network

3.SVM

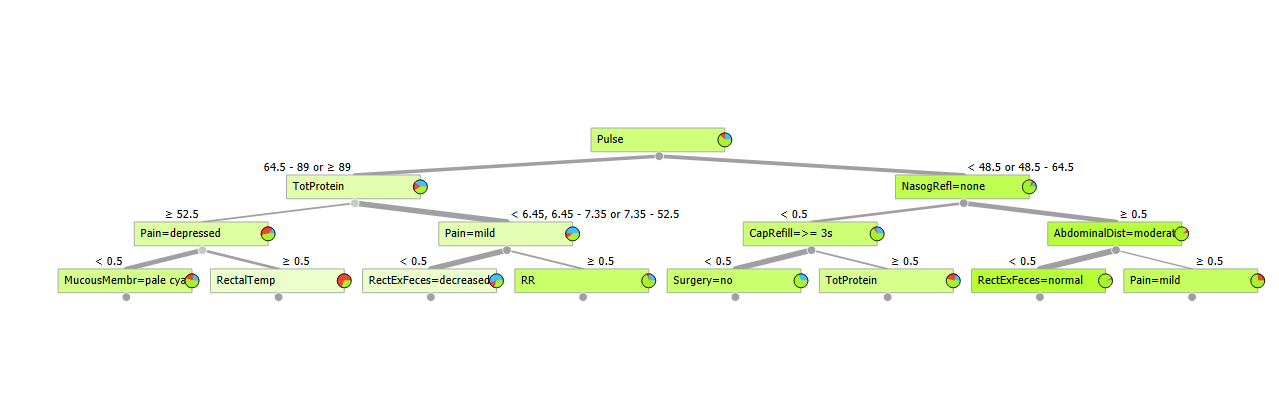
4.Navie Bayes

5.Tree

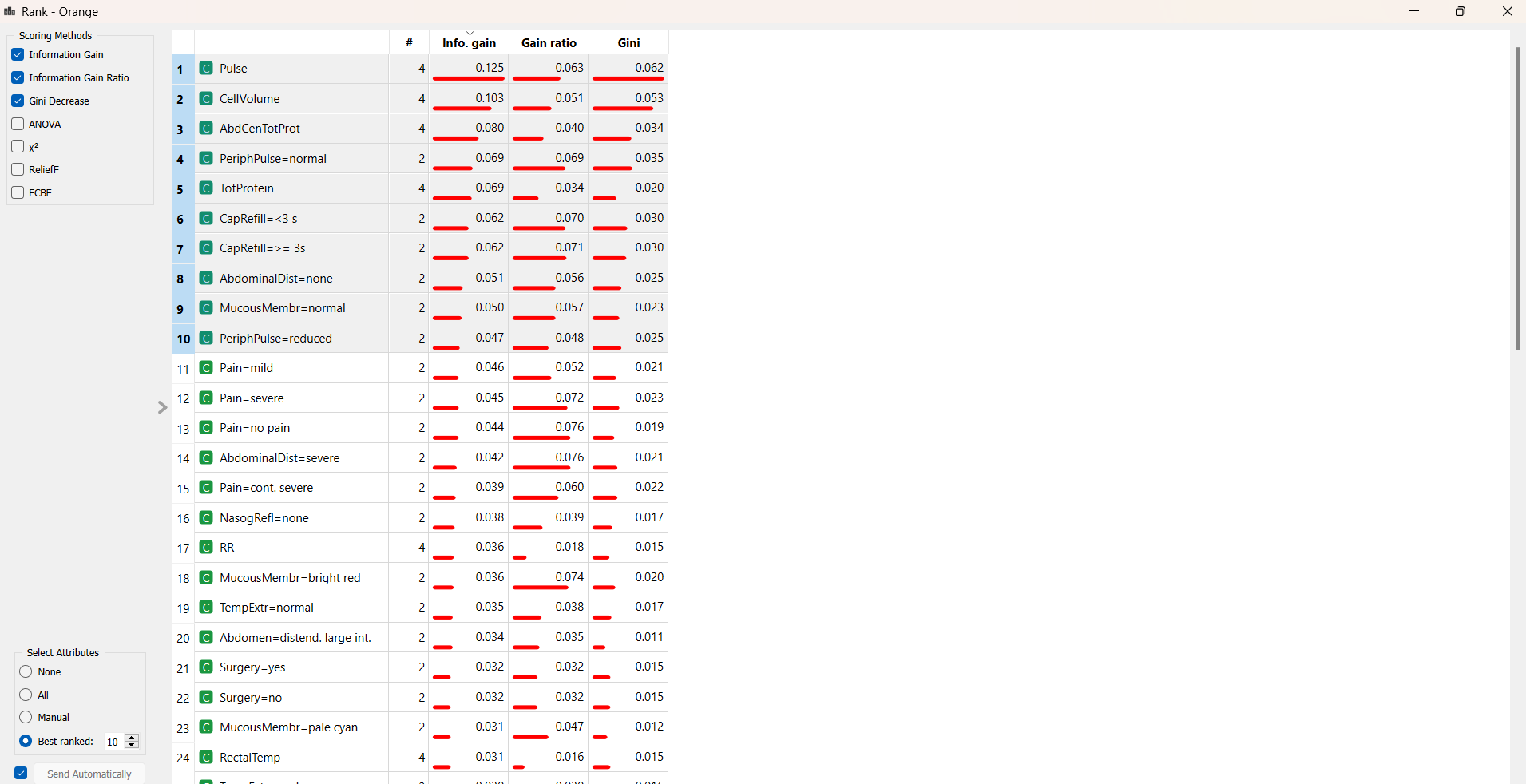
TEST AND SCORE:



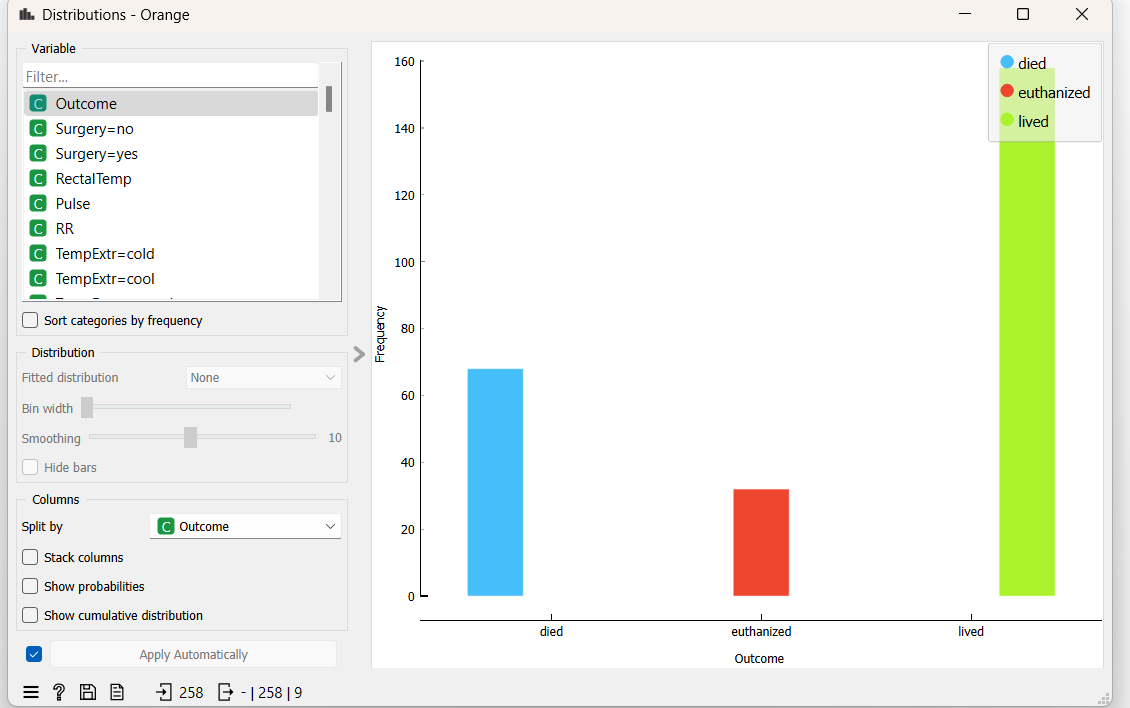
TREE:

Pulse is declared as root because the information gain of pulse is highest of all.

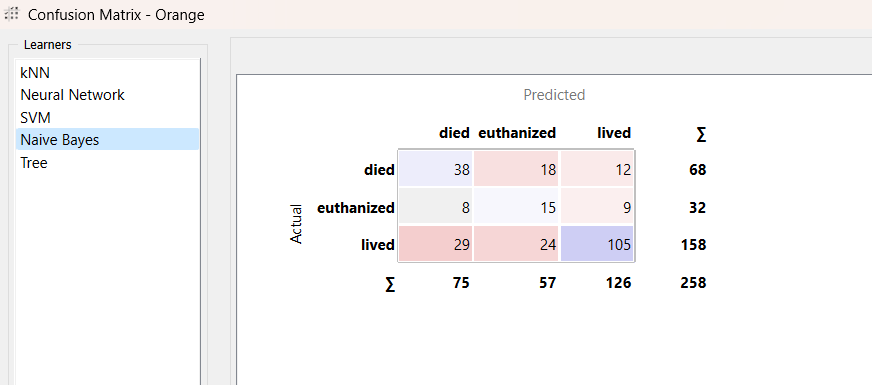
RANK :



DISTRIBUTIONS:



CONFUSION MATRIX:



OBSERVATIONS:

1.Total 21 attributes(input:20,output:1)

2.Preprocessing: Impute missing values

Continuous discrete variables

3.Classification techniques used: KNN ,Neural Network ,SVM ,Navie Bayes ,Tree

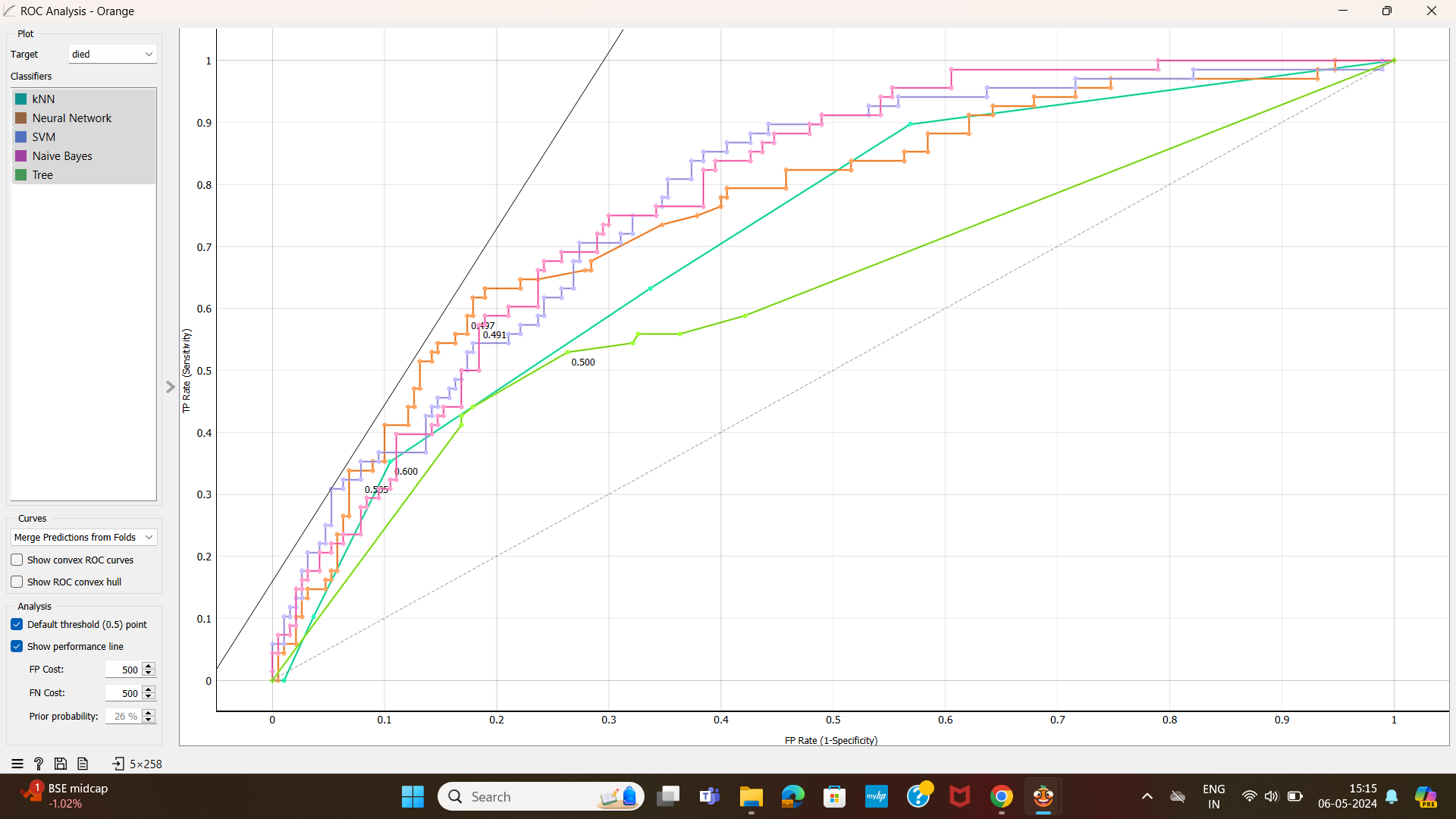
4.In tree we find “pulse” as root

5.The values of Accuracy ,Precision ,F1\_score are nearer to each other in confusion matrix

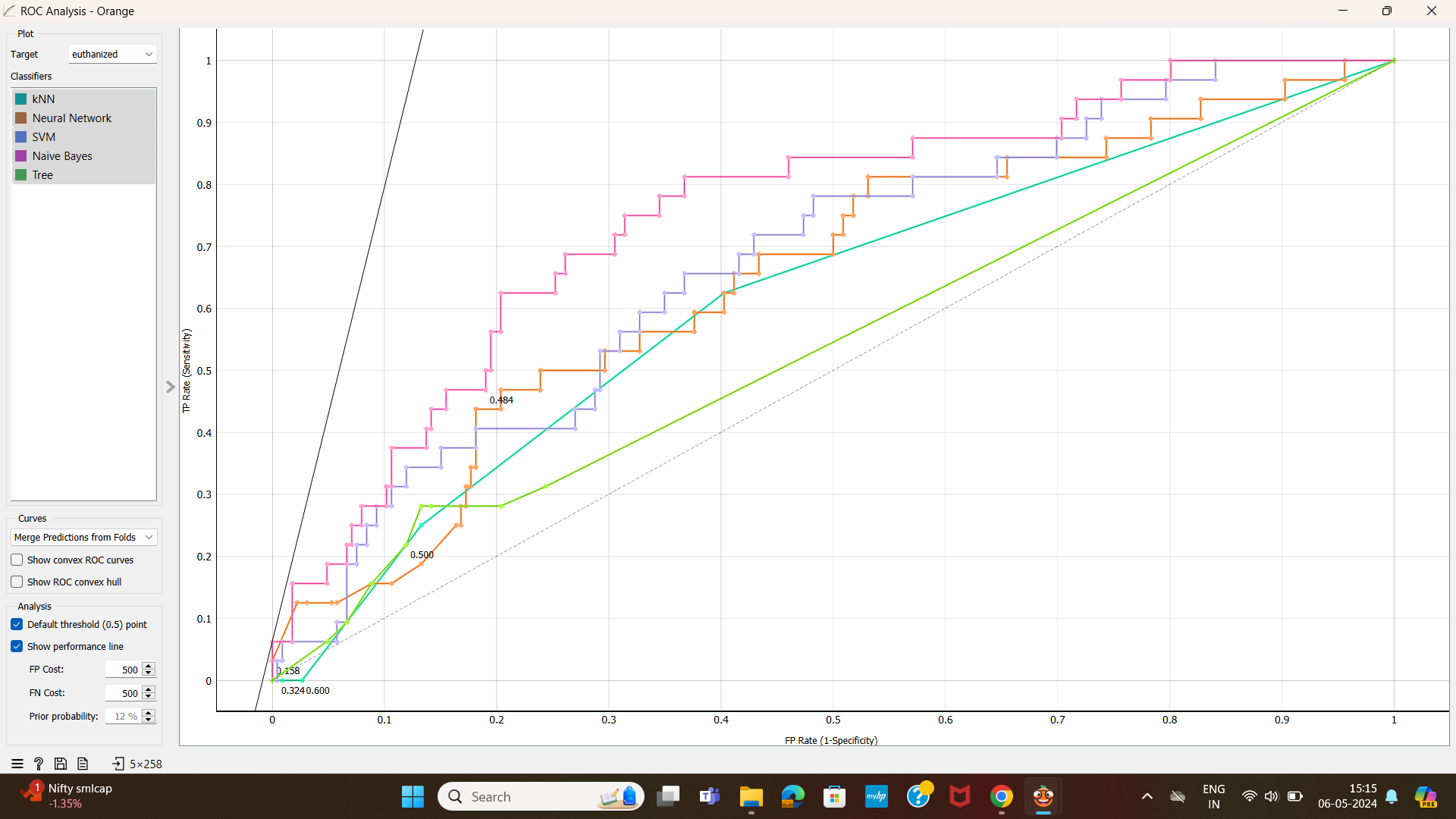
OUTPUT:

ROC Analysis

For target : died



For target:enthanized



For target : lived

