SPRAWOZDANIE

Zajęcia: Uczenie Maszynowe

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 2 Data 09.11.2024 Temat: "Praktyczne Zastosowanie Drzew Decyzyjnych i Metod Ensemble w Analizie Danych" Wariant 11	Szymon Nycz Informatyka II stopień, niestacjonarne, 1 semestr, gr.1b

1. Polecenie:

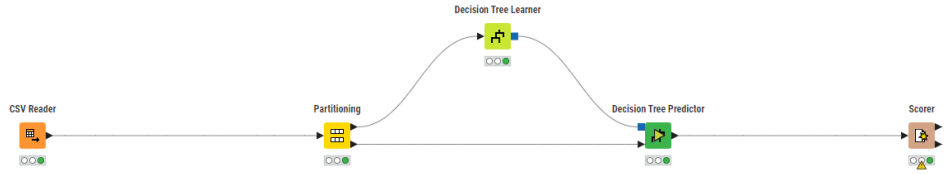
   Powikłania zawału mięśnia sercowego:
   https://www.kaggle.com/datasets/rafatashrafjoy/myocardial-infarction-complications

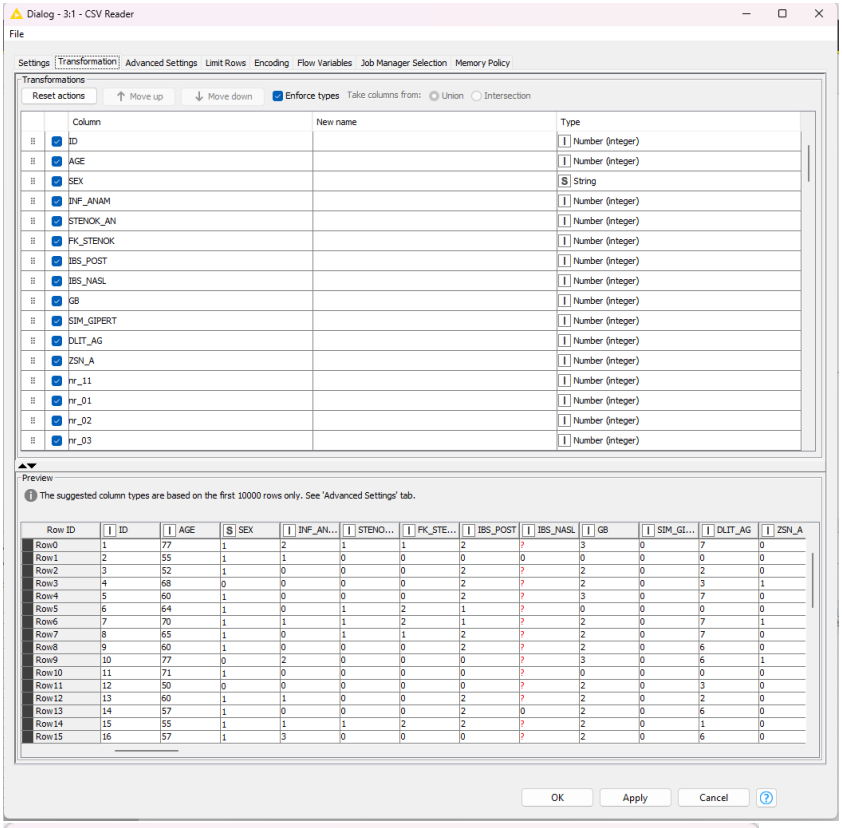
2. Link do repozytorium:

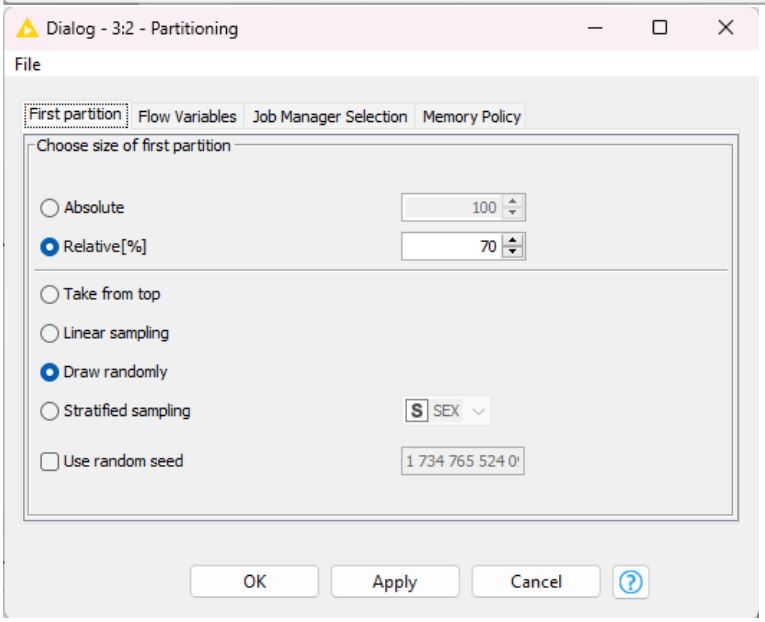
   Link: https://github.com/Maciek332/Semestr_1_Nycz/tree/master/UM

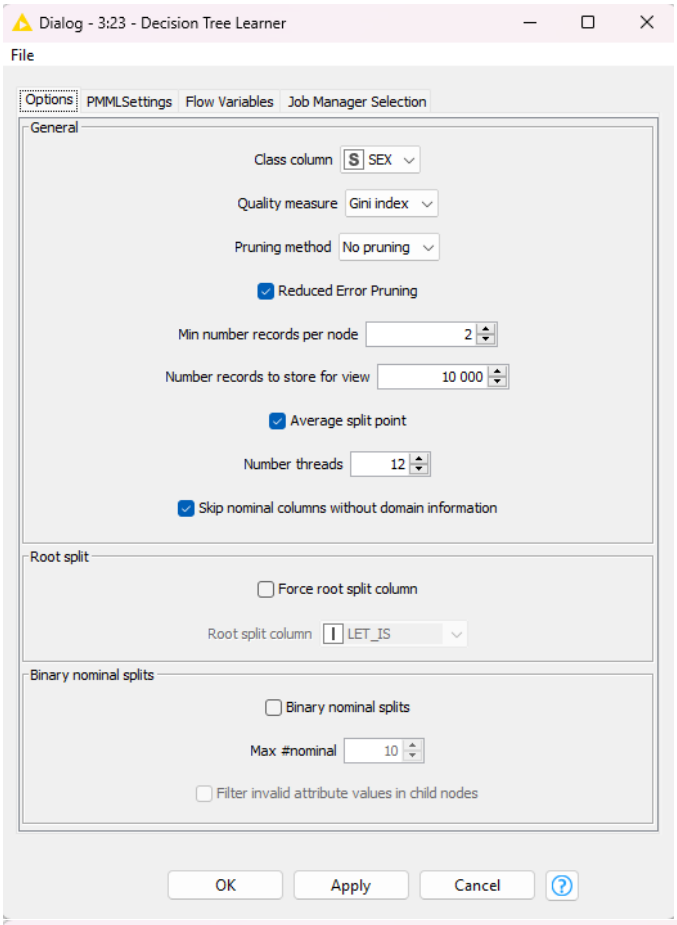
3. Opis programu opracowanego

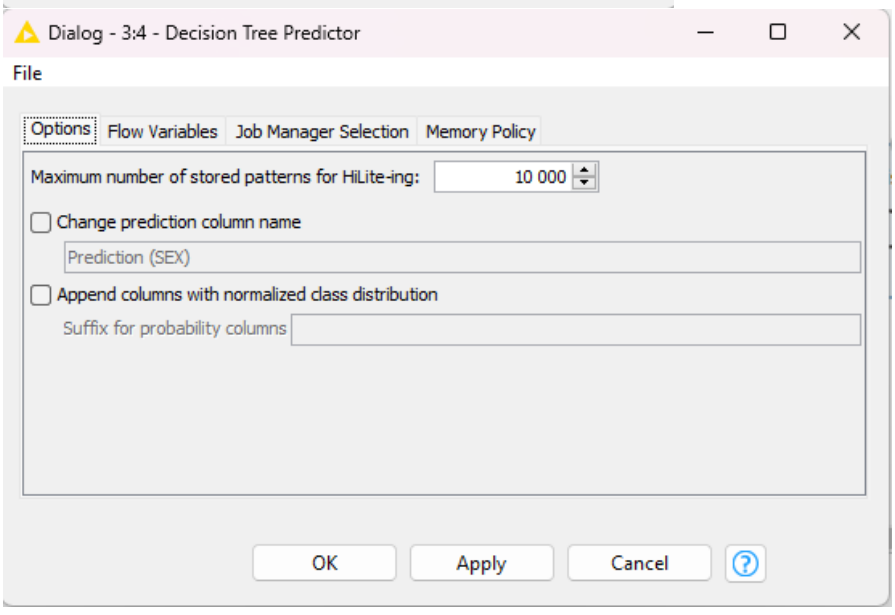
- Decyzjonalne drzewo przepływów

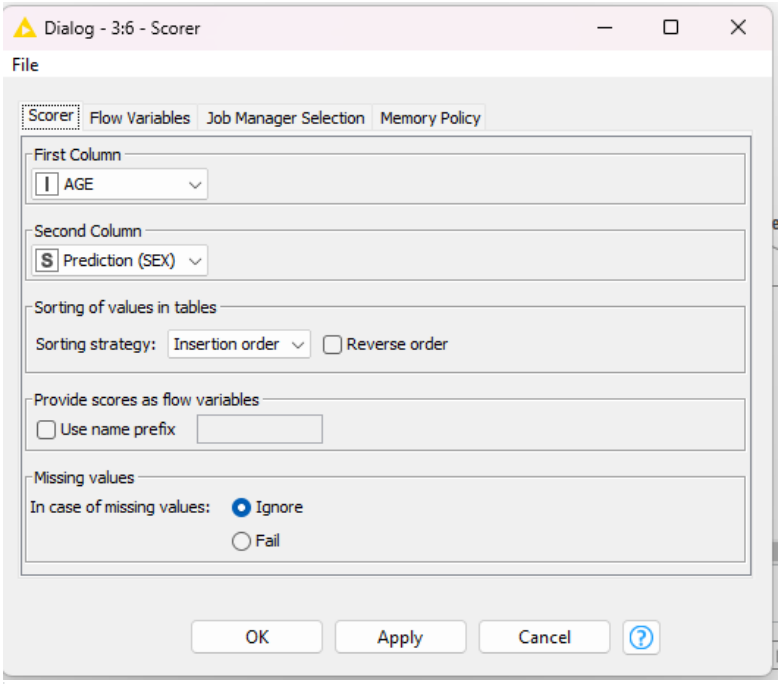




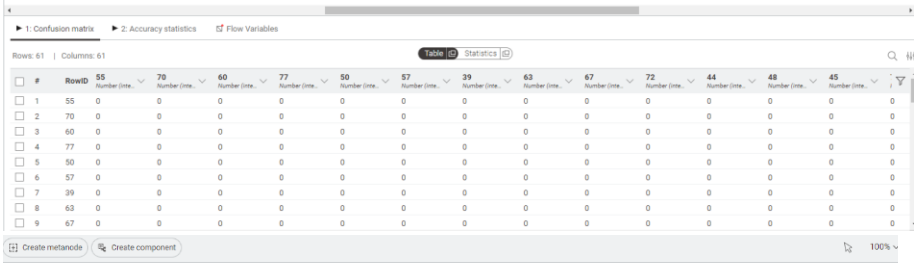



Dialog - 3:23 - Decision Tree Learner
File
Options
PMMLSettings
Flow Variables
Job Manager Selection
General
Class column
S SEX
Quality measure
Gini index
Pruning method
No pruning
Reduced Error Pruning
Min number records per node
2
Number records to store for view
10 000
Average split point
Number threads
12
Skip nominal columns without domain information
Root split
Force root split column
Root split column
LET_IS
Binary nominal splits
Binary nominal splits
Max #nominal
10
Filter invalid attribute values in child nodes
OK
Apply
Cancel

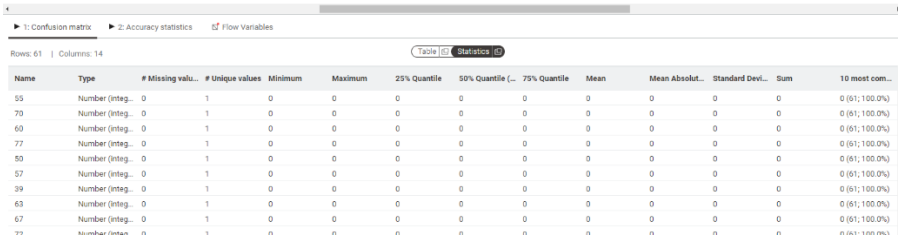

Dialog - 3:4 - Decision Tree Predictor
File
Options
Flow Variables
Job Manager Selection
Memory Policy
Maximum number of stored patterns for HiLite-ing:
10 000
Change prediction column name
Prediction (SEX)
Append columns with normalized class distribution
Suffix for probability columns
OK
Apply
Cancel


Dialog - 3:6 - Scorer
File
Scorer
Flow Variables
Job Manager Selection
Memory Policy
First Column
AGE
Second Column
Prediction (SEX)
Sorting of values in tables
Sorting strategy:
Insertion order
Reverse order
Provide scores as flow variables
Use name prefix
Missing values
In case of missing values:
Ignore
Fail
OK
Apply
Cancel

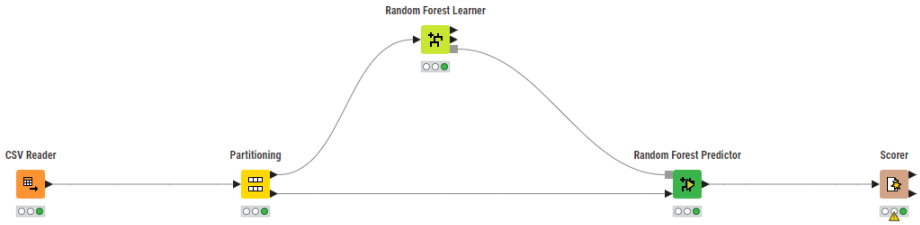

Decision Tree Learner
CSV Reader
Partitioning
Decision Tree Predictor
Scorer
Add comment

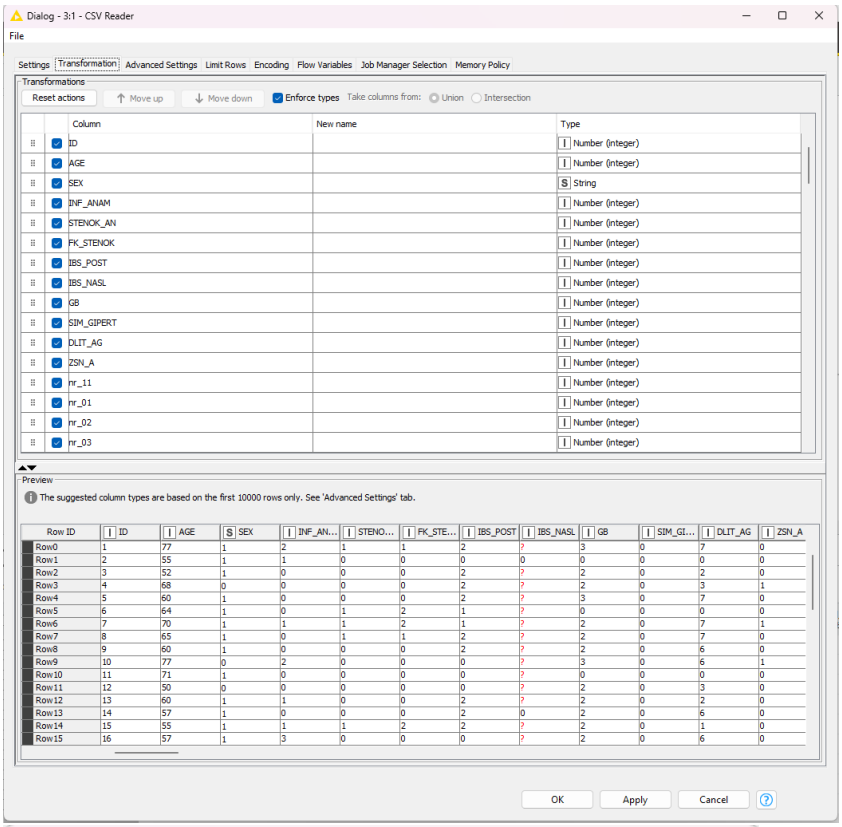

1: Confusion matrix
2: Accuracy statistics
Flow Variables
Rows: 61 | Columns: 61
Table
Statistics
Create metanode
Create component

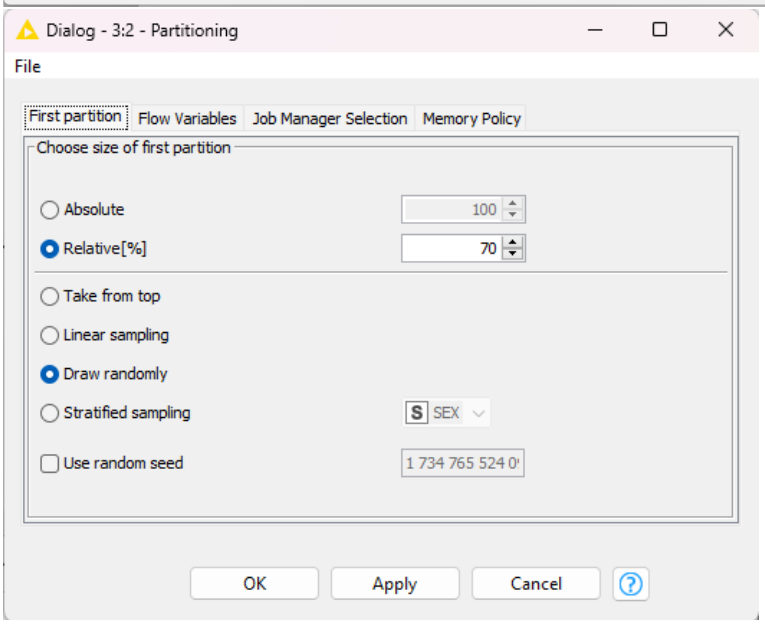

Decision Tree Learner
CSV Reader
Partitioning
Decision Tree Predictor
Scorer
Add comment

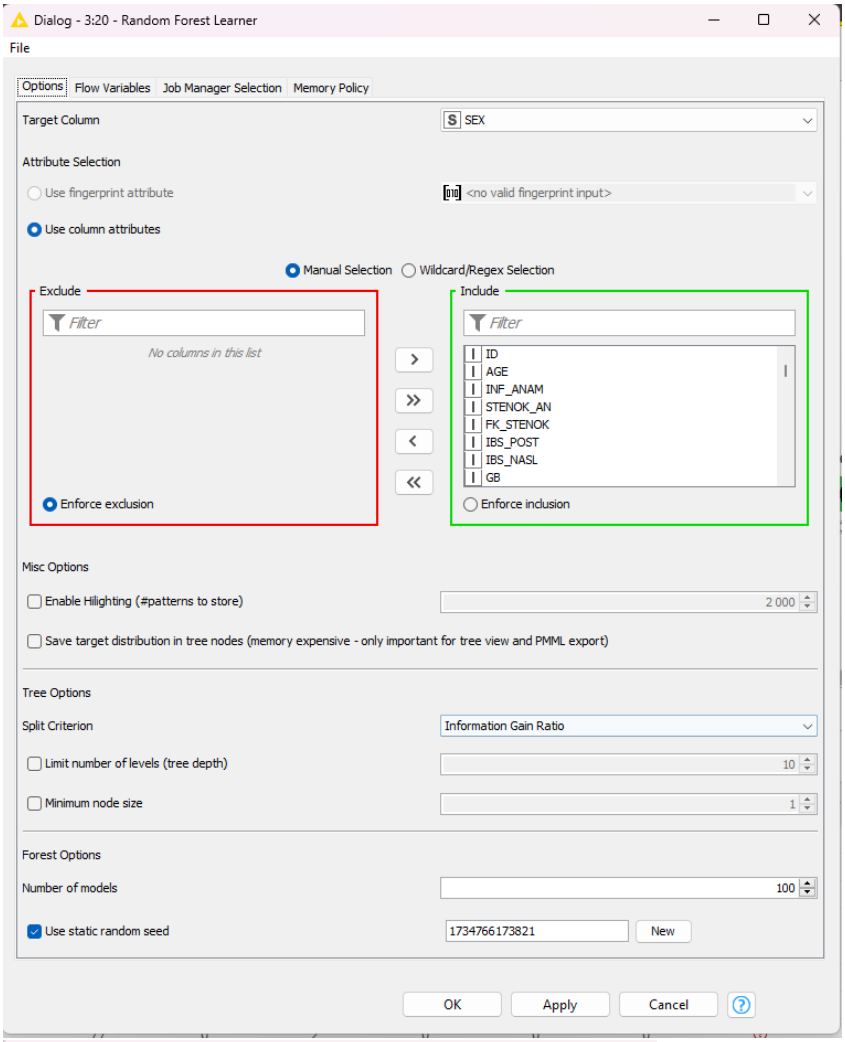

1: Confusion matrix
2: Accuracy statistics
Flow Variables
Rows: 61 | Columns: 14
Table
Statistics

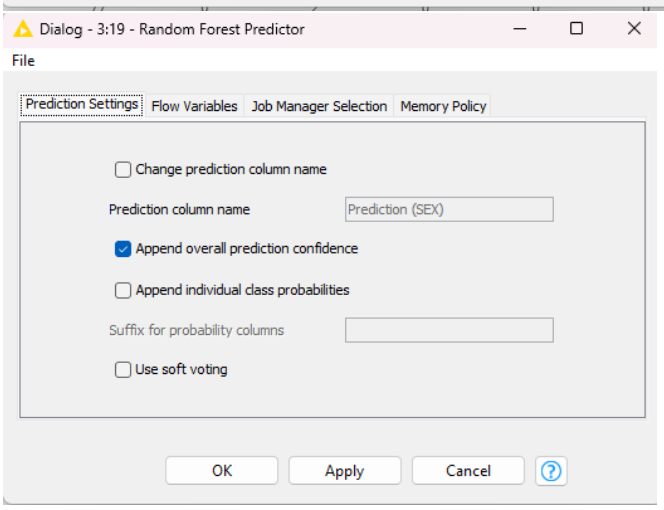
- Las losowy

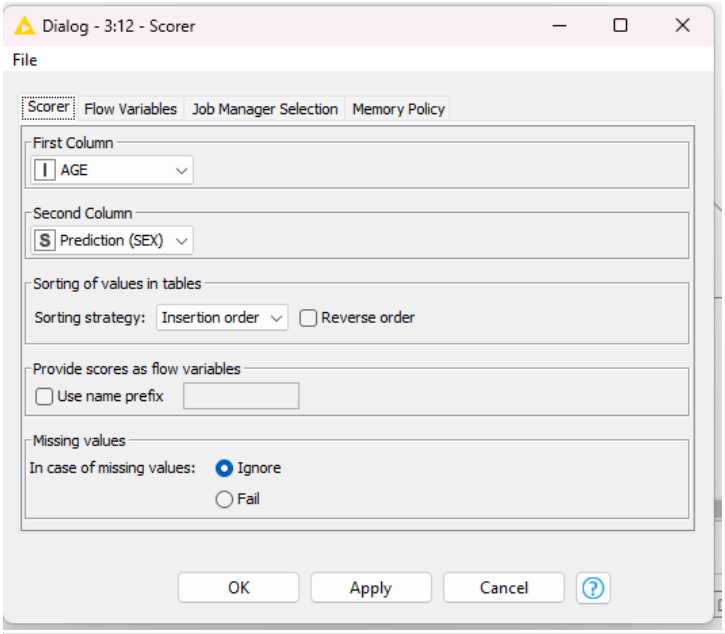

Random Forest Learner
CSV Reader
Partitioning
Random Forest Predictor
Scorer

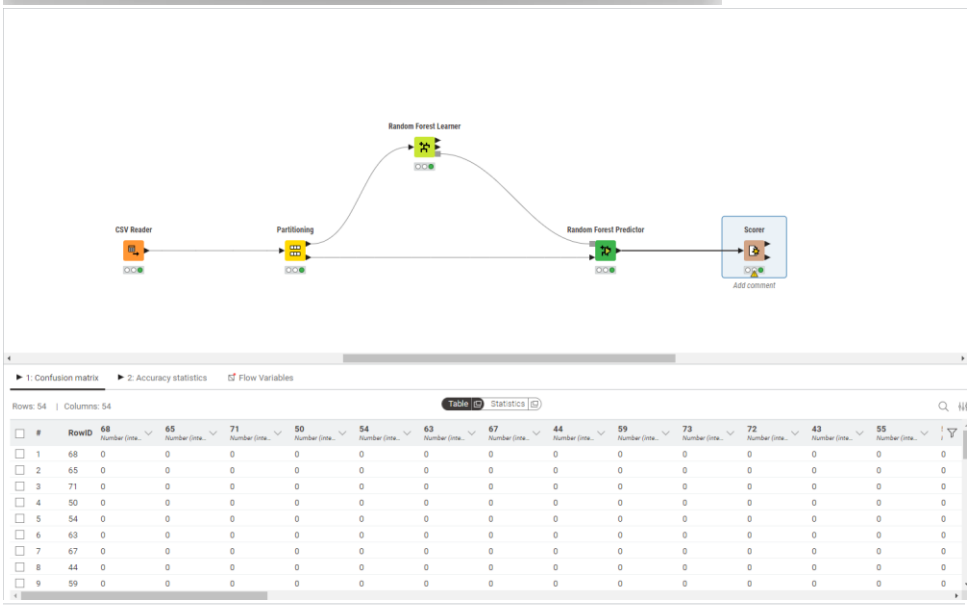

Dialog - 3:1 - CSV Reader
File
Settings
Transformation
Advanced Settings
Limit Rows
Encoding
Flow Variables
Job Manager Selection
Memory Policy
Transformations
Reset actions
Move up
Move down
Enforce types
Take columns from:
Union
Intersection
Column
New name
Type
ID
AGE
SEX
INF_ANAM
STENOK_AN
FK_STENOK
IBS_POST
IBS_NASL
GB
SIM_GIPERT
DLIT_AG
ZSN_A
nr_11
nr_01
nr_02
nr_03
Number (integer)
String
Preview
The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.
OK
Apply
Cancel

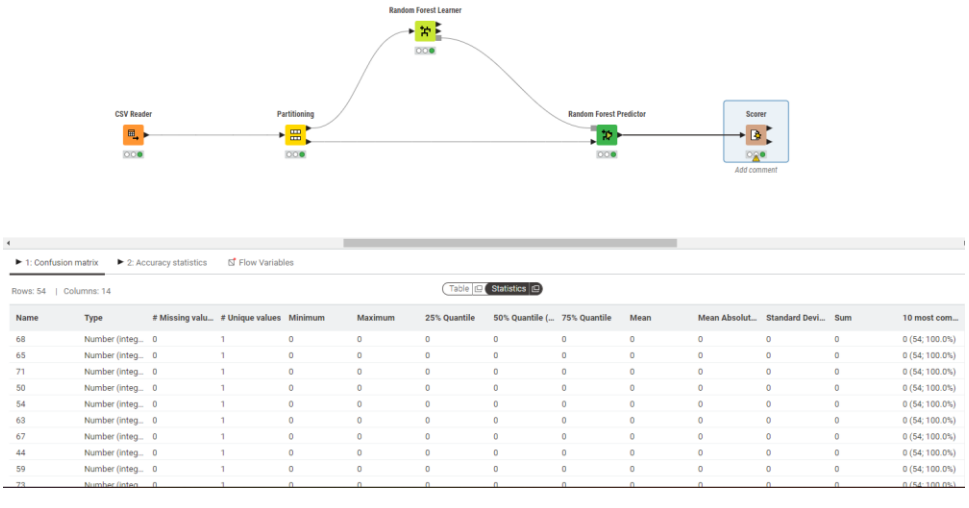

Dialog - 3:2 - Partitioning
File
First partition
Flow Variables
Job Manager Selection
Memory Policy
Choose size of first partition
Absolute
100
Relative[%]
70
Take from top
Linear sampling
Draw randomly
Stratified sampling
SEX
Use random seed
1 734 765 524 0
OK
Apply
Cancel


Dialog - 3:20 - Random Forest Learner
File
Options
Flow Variables
Job Manager Selection
Memory Policy
Target Column
SEX
Attribute Selection
Use fingerprint attribute
<no valid fingerprint input>
Use column attributes
Manual Selection
Wildcard/Regex Selection
Exclude
Filter
No columns in this list
Enforce exclusion
Include
Filter
ID
AGE
INF_ANAM
STENOK_AN
FK_STENOK
IBS_POST
IBS_NASL
GB
Enforce inclusion
Misc Options
Enable Hilighting (#patterns to store)
2 000
Save target distribution in tree nodes (memory expensive - only important for tree view and PMML export)
Tree Options
Split Criterion
Information Gain Ratio
Limit number of levels (tree depth)
10
Minimum node size
1
Forest Options
Number of models
100
Use static random seed
1734766173821
New
OK
Apply
Cancel

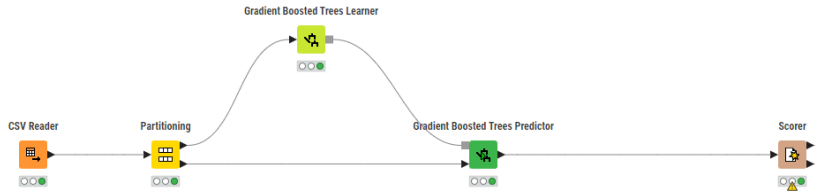

Dialog - 3:19 - Random Forest Predictor
File
Prediction Settings
Flow Variables
Job Manager Selection
Memory Policy
Change prediction column name
Prediction column name
Prediction (SEX)
Append overall prediction confidence
Append individual class probabilities
Suffix for probability columns
Use soft voting
OK
Apply
Cancel

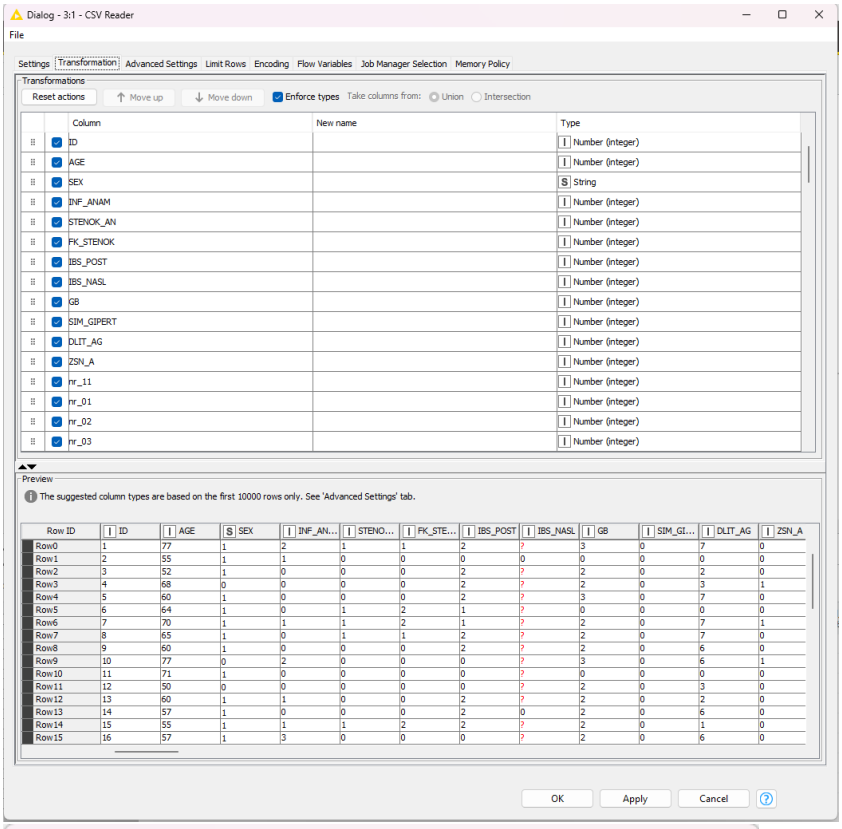

Dialog - 3:12 - Scorer
File
Scorer
Flow Variables
Job Manager Selection
Memory Policy
First Column
AGE
Second Column
Prediction (SEX)
Sorting of values in tables
Sorting strategy:
Insertion order
Reverse order
Provide scores as flow variables
Use name prefix
Missing values
In case of missing values:
Ignore
Fail
OK
Apply
Cancel

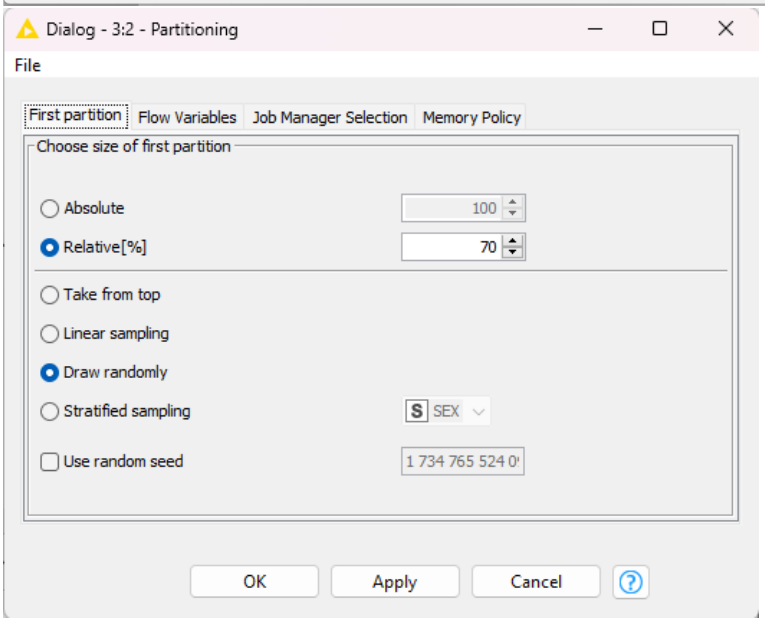

Random Forest Learner
CSV Reader
Partitioning
Random Forest Predictor
Scorer
Add comment
1: Confusion matrix
2: Accuracy statistics
Flow Variables
Rows: 54 | Columns: 54
Table
Statistics

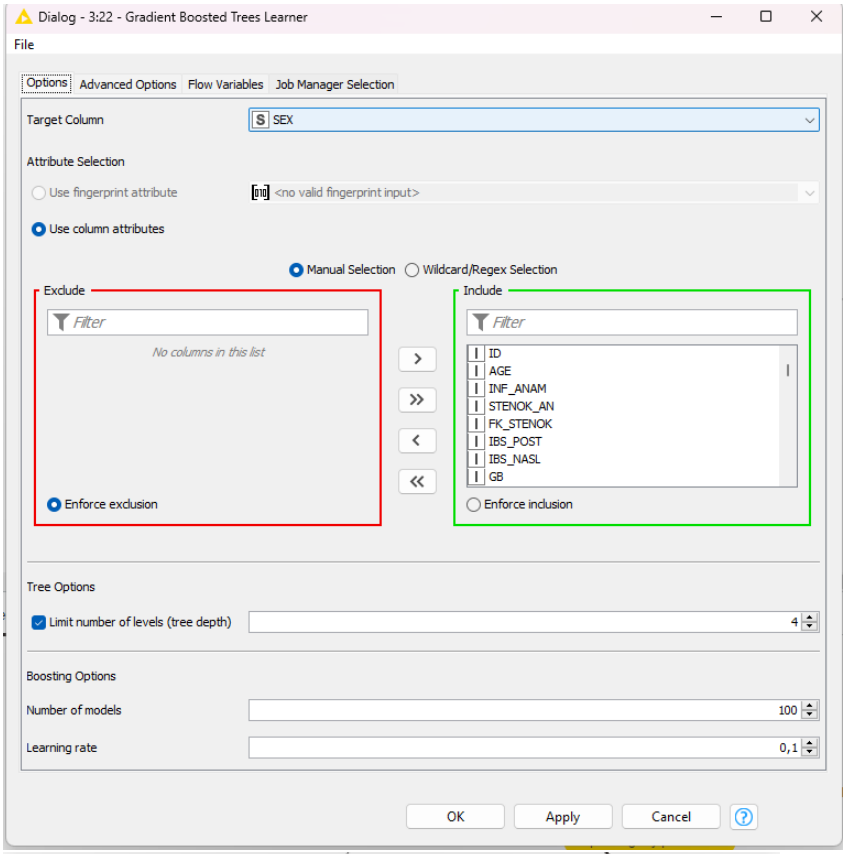

Random Forest Learner
CSV Reader
Partitioning
Random Forest Predictor
Scorer
Add comment
1: Confusion matrix
2: Accuracy statistics
Flow Variables
Rows: 54 | Columns: 14
Table
Statistics

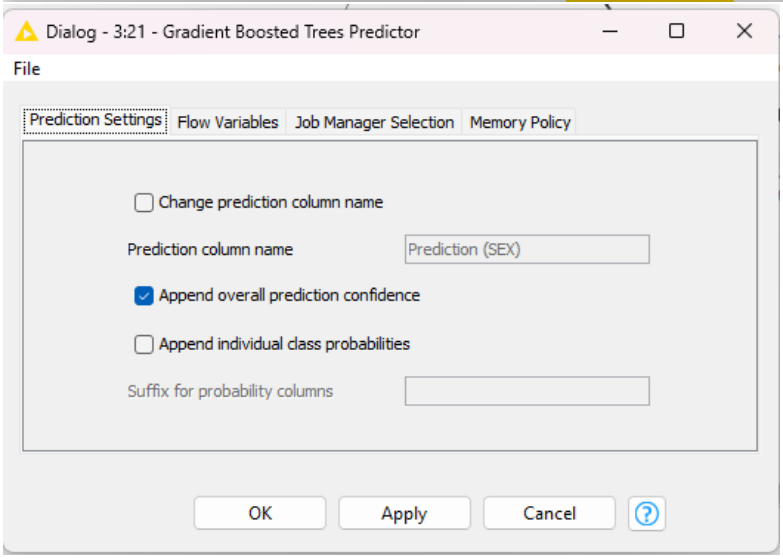
- Boosted trees

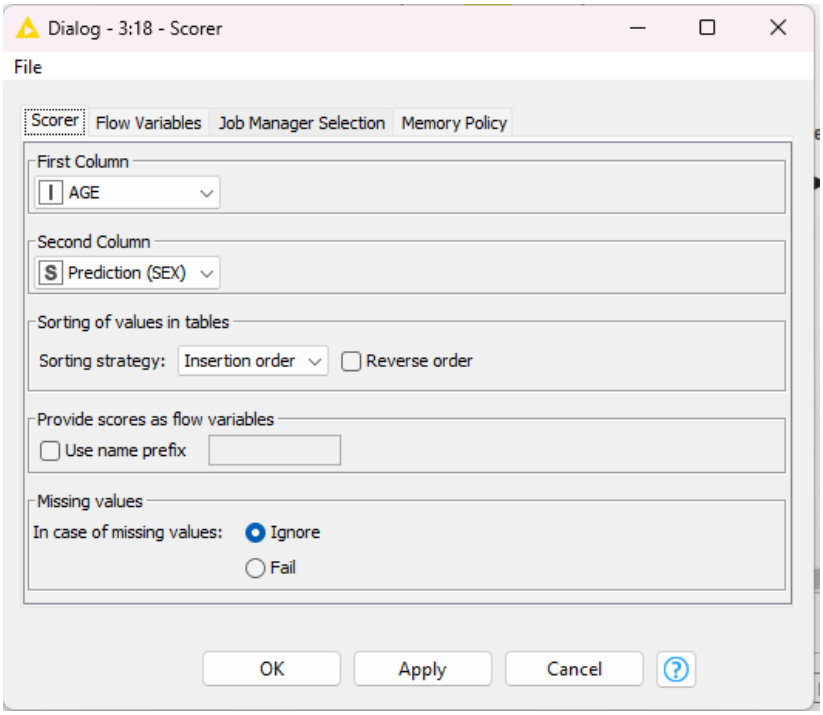


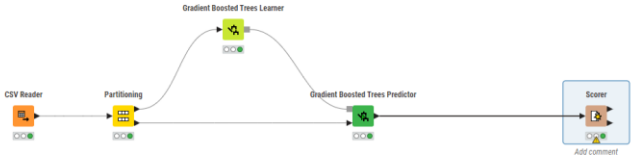


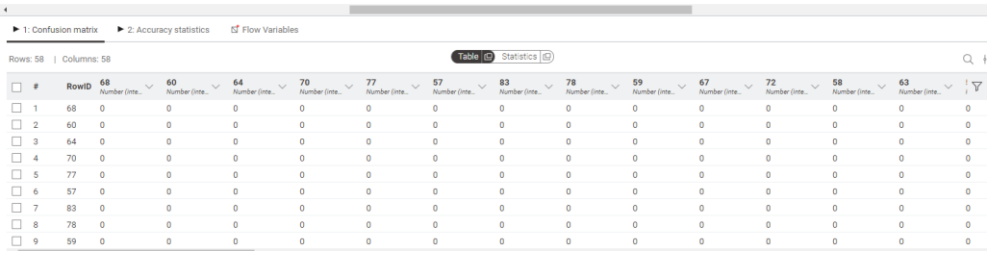


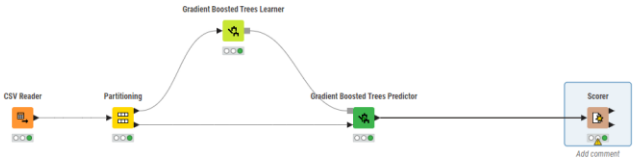

Dialog - 3:22 - Gradient Boosted Trees Learner
File
Options
Advanced Options
Flow Variables
Job Manager Selection
Target Column
SEX
Attribute Selection
Use fingerprint attribute
<no valid fingerprint input>
Use column attributes
Manual Selection
Wildcard/Regex Selection
Exclude
Filter
No columns in this list
Enforce exclusion
Include
Filter
ID
AGE
INF_ANAM
STENOK_AN
FK_STENOK
IBS_POST
IBS_NASL
GB
Enforce inclusion
Tree Options
Limit number of levels (tree depth)
4
Boosting Options
Number of models
100
Learning rate
0,1
OK
Apply
Cancel

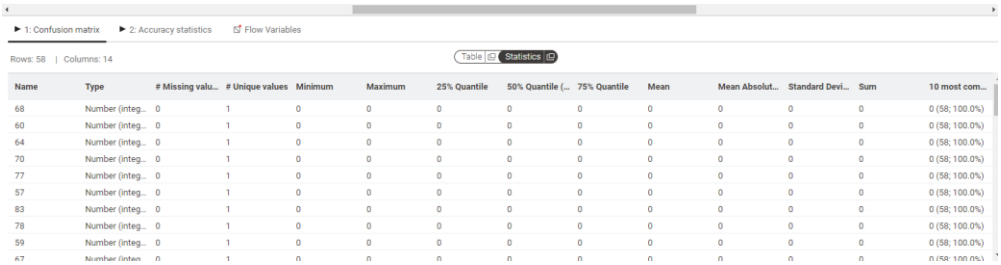

Dialog - 3:21 - Gradient Boosted Trees Predictor
File
Prediction Settings
Flow Variables
Job Manager Selection
Memory Policy
Change prediction column name
Prediction column name
Prediction (SEX)
Append overall prediction confidence
Append individual class probabilities
Suffix for probability columns
OK
Apply
Cancel


Dialog - 3:18 - Scorer
File
Scorer
Flow Variables
Job Manager Selection
Memory Policy
First Column
AGE
Second Column
Prediction (SEX)
Sorting of values in tables
Sorting strategy:
Insertion order
Reverse order
Provide scores as flow variables
Use name prefix
Missing values
In case of missing values:
Ignore
Fail
OK
Apply
Cancel


Gradient Boosted Trees Learner
CSV Reader
Partitioning
Gradient Boosted Trees Predictor
Scorer
Add comment


1: Confusion matrix
2: Accuracy statistics
Flow Variables
Rows: 58
Columns: 58
Table
Statistics





4. Wnioski

KNIME oferuje efektywne i intuicyjne narzędzie do tworzenia modeli klasyfikacyjnych, wykorzystujące graficzne przepływy pracy. Dzięki węzłom KNIME, można łatwo wdrożyć modele takie jak drzewa decyzyjne, Random Forest i boosting, umożliwiając analizę danych bez potrzeby kodowania. Drzewa decyzyjne są prostymi, ale skutecznymi modelami uczenia maszynowego, szczególnie przydatnymi w analizie danych. Metody zbiorcze, takie jak bagging, Random Forest i boosting, zwiększają dokładność i stabilność modeli poprzez łączenie wielu słabszych klasyfikatorów. Random Forest dodatkowo wprowadza element losowości przy wyborze cech.