

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:

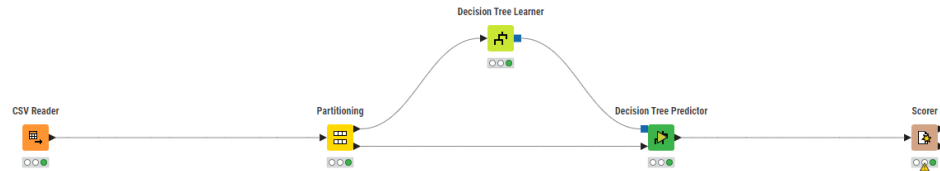
<https://www.kaggle.com/datasets/rafatashrafjoy/myocardial-infarction-complications>

2. Link do repozytorium:

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

3. Opis programu opracowanego

- Decyzjonalne drzewo przepływów



Dialog - 3:1 - 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
<input checked="" type="checkbox"/> ID		I Number (integer)
<input checked="" type="checkbox"/> AGE		I Number (integer)
<input checked="" type="checkbox"/> SEX		S String
<input checked="" type="checkbox"/> INF_AMAM		I Number (integer)
<input checked="" type="checkbox"/> STENO_K_AN		I Number (integer)
<input checked="" type="checkbox"/> FK_STENO_K		I Number (integer)
<input checked="" type="checkbox"/> IBS_POST		I Number (integer)
<input checked="" type="checkbox"/> IBS_NASL		I Number (integer)
<input checked="" type="checkbox"/> GB		I Number (integer)
<input checked="" type="checkbox"/> SIM_GIPERT		I Number (integer)
<input checked="" type="checkbox"/> DLT_AG		I Number (integer)
<input checked="" type="checkbox"/> ZSN_A		I Number (integer)
<input checked="" type="checkbox"/> nr_11		I Number (integer)
<input checked="" type="checkbox"/> nr_01		I Number (integer)
<input checked="" type="checkbox"/> nr_02		I Number (integer)
<input checked="" type="checkbox"/> nr_03		I Number (integer)

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	ID	AGE	SEX	INF_AM...	STENO...	FK_STE...	IBS_POST	IBS_NASL	GB	SIM_GL...	DLT_AG	ZSN_A
Row0	1	77	1	2	1	1	2	?	3	0	7	0
Row1	2	55	1	1	0	0	0	?	0	0	0	0
Row2	3	52	1	0	0	0	2	?	2	0	2	0
Row3	4	68	0	0	0	0	2	?	2	0	3	1
Row4	5	60	1	0	0	0	2	?	3	0	7	0
Row5	6	64	1	0	1	2	1	?	0	0	0	0
Row6	7	70	1	1	1	2	1	?	2	0	7	1
Row7	8	65	1	0	1	1	2	?	2	0	7	0
Row8	9	60	1	0	0	0	2	?	2	0	6	0
Row9	10	77	0	2	0	0	0	?	3	0	6	1
Row10	11	71	1	0	0	0	0	?	0	0	0	0
Row11	12	50	0	0	0	0	0	?	2	0	3	0
Row12	13	60	1	1	0	0	2	?	2	0	2	0
Row13	14	57	1	0	0	0	2	?	2	0	6	0
Row14	15	55	1	1	1	2	2	?	2	0	1	0
Row15	16	57	1	3	0	0	0	?	2	0	6	0

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 S SEX

☐ Use random seed 1 734 765 524 0

OK Apply Cancel ?

Dialog - 3:23 - Decision Tree Learner

File

Options PMMLSettings Flow Variables Job Manager Selection

General

Class column

Quality measure

Pruning method

☒ Reduced Error Pruning

Min number records per node

Number records to store for view

☒ Average split point

Number threads

☒ Skip nominal columns without domain information

Root split

☐ Force root split column

Root split column

Binary nominal splits

☐ Binary nominal splits

Max #nominal

☐ 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:

☐ Change prediction column name

☐ Append columns with normalized class distribution

Suffix for probability columns

OK Apply Cancel ?

The screenshot displays the Orange3 data mining software interface. At the top, a workflow is visualized with the following components and connections:

- CSV Reader** (orange icon) connects to **Partitioning** (yellow icon).
- Partitioning** connects to **Decision Tree Learner** (green icon).
- Decision Tree Learner** connects to **Decision Tree Predictor** (green icon).
- Decision Tree Predictor** connects to **Scorer** (blue icon).
- A text label "Add comment" is positioned below the Scorer widget.

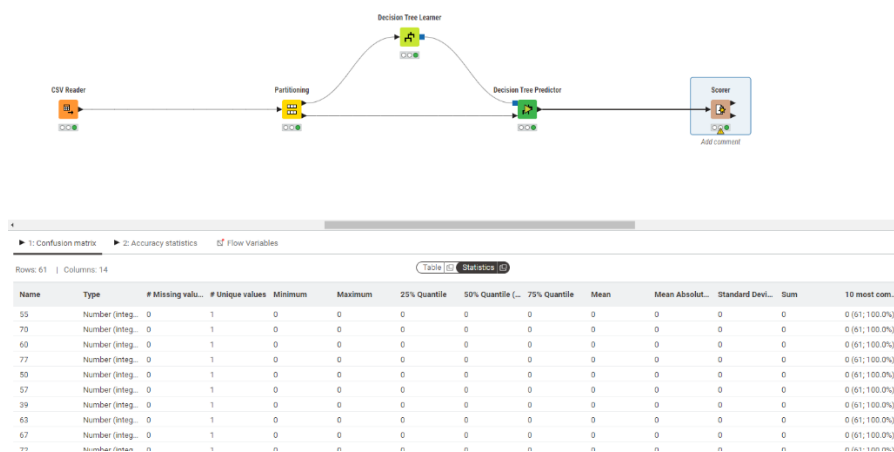
Below the workflow, the interface shows a summary of the data and a table of the first 10 rows:

- Summary: Rows: 61 | Columns: 61
- Navigation tabs: 1: Confusion matrix, 2: Accuracy statistics, Flow Variables.
- Table tabs: Table (selected), Statistics.

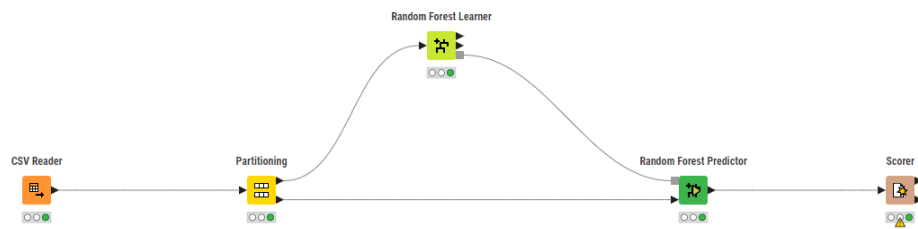
The table has the following structure:

#	RowID	55	70	60	77	50	57	39	63	67	72	44	48	45
		Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)	Number (int.)
1	55	0	0	0	0	0	0	0	0	0	0	0	0	0
2	70	0	0	0	0	0	0	0	0	0	0	0	0	0
3	60	0	0	0	0	0	0	0	0	0	0	0	0	0
4	77	0	0	0	0	0	0	0	0	0	0	0	0	0
5	50	0	0	0	0	0	0	0	0	0	0	0	0	0
6	57	0	0	0	0	0	0	0	0	0	0	0	0	0
7	39	0	0	0	0	0	0	0	0	0	0	0	0	0
8	63	0	0	0	0	0	0	0	0	0	0	0	0	0
9	67	0	0	0	0	0	0	0	0	0	0	0	0	0

At the bottom, there are buttons for "Create metaclass" and "Create component", and a status bar showing "100%" zoom level.



- Las losowy



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
<input checked="" type="checkbox"/>	ID		I Number (integer)
<input checked="" type="checkbox"/>	AGE		I Number (integer)
<input checked="" type="checkbox"/>	SEX		S String
<input checked="" type="checkbox"/>	INF_ANAM		I Number (integer)
<input checked="" type="checkbox"/>	STENO_KOK_AN		I Number (integer)
<input checked="" type="checkbox"/>	FK_STENO_KOK		I Number (integer)
<input checked="" type="checkbox"/>	IBS_POST		I Number (integer)
<input checked="" type="checkbox"/>	IBS_NASL		I Number (integer)
<input checked="" type="checkbox"/>	GB		I Number (integer)
<input checked="" type="checkbox"/>	SIM_GPERT		I Number (integer)
<input checked="" type="checkbox"/>	DLIT_AG		I Number (integer)
<input checked="" type="checkbox"/>	ZSN_A		I Number (integer)
<input checked="" type="checkbox"/>	nr_11		I Number (integer)
<input checked="" type="checkbox"/>	nr_01		I Number (integer)
<input checked="" type="checkbox"/>	nr_02		I Number (integer)
<input checked="" type="checkbox"/>	nr_03		I Number (integer)

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	T ID	T AGE	S SEX	T INF_AN...	T STENO...	T FK_STE...	T IBS_POST	T IBS_NASL	T GB	T SIM_GL...	T DLIT_AG	T ZSN_A
Row0	1	77	1	2	1	2	0	3	0	7	0	
Row1	2	55	1	1	0	0	0	0	0	0	0	
Row2	3	52	1	0	0	0	2	2	0	2	0	
Row3	4	68	0	0	0	0	2	2	0	3	1	
Row4	5	60	1	0	0	0	2	3	0	7	0	
Row5	6	64	1	0	1	2	1	0	0	0	0	
Row6	7	70	1	1	1	2	1	2	0	7	1	
Row7	8	65	1	0	1	1	2	2	0	7	0	
Row8	9	60	1	0	0	0	2	2	0	6	0	
Row9	10	77	0	2	0	0	0	3	0	6	1	
Row10	11	71	1	0	0	0	0	0	0	0	0	
Row11	12	50	0	0	0	0	0	2	0	3	0	
Row12	13	60	1	1	0	0	2	2	0	2	0	
Row13	14	57	1	0	0	0	2	0	0	6	0	
Row14	15	55	1	1	1	2	2	2	0	1	0	
Row15	16	57	1	3	0	0	0	2	0	6	0	

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 S 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

☒ 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 Hlilghting (#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

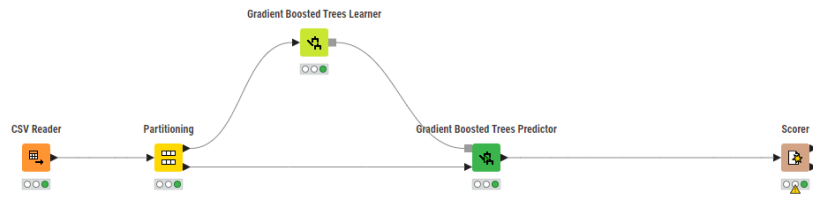
☐ Append individual class probabilities

Suffix for probability columns:

☐ Use soft voting

OK Apply Cancel ?

- Boosted trees



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		Number (integer)
AGE		Number (integer)
SEX		String
INF_ANAM		Number (integer)
STENOK_AN		Number (integer)
FK_STENOK		Number (integer)
IBS_POST		Number (integer)
IBS_NASL		Number (integer)
GB		Number (integer)
SIM_GPERT		Number (integer)
DLIT_AG		Number (integer)
ZSN_A		Number (integer)
nr_11		Number (integer)
nr_01		Number (integer)
nr_02		Number (integer)
nr_03		Number (integer)

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	ID	AGE	SEX	INF_ANAM	STENO...	FK_STE...	IBS_POST	IBS_NASL	GB	SIM_GL...	DLIT_AG	ZSN_A
Row0	1	77	1	2	1	1	2	3	0	7	0	
Row1	2	55	1	1	0	0	0	0	0	0	0	
Row2	3	52	1	0	0	0	2	2	0	2	0	
Row3	4	68	0	0	0	0	2	2	0	3	1	
Row4	5	60	1	0	0	0	2	3	0	7	0	
Row5	6	64	1	0	1	2	1	0	0	0	0	
Row6	7	70	1	1	1	2	1	2	0	7	1	
Row7	8	65	1	0	1	1	2	2	0	7	0	
Row8	9	60	1	0	0	0	2	2	0	6	0	
Row9	10	77	0	2	0	0	0	3	0	6	1	
Row10	11	71	1	0	0	0	0	0	0	0	0	
Row11	12	50	0	0	0	0	0	2	0	3	0	
Row12	13	60	1	1	0	0	2	2	0	2	0	
Row13	14	57	1	0	0	0	2	0	2	6	0	
Row14	15	55	1	1	1	2	2	2	0	1	0	
Row15	16	57	1	3	0	0	0	2	0	6	0	

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 S SEX

☐ Use random seed 1 734 765 524 0

OK Apply Cancel ?

Dialog - 3:22 - Gradient Boosted Trees Learner

File

Options Advanced Options Flow Variables Job Manager Selection

Target Column

Attribute Selection

☐ Use fingerprint attribute

☒ Use column attributes

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

No columns in this list

☒ Enforce exclusion

Include

- ☐ ID
- ☐ AGE
- ☐ INF_ANAM
- ☐ STENOK_AN
- ☐ FK_STENOK
- ☐ IBS_POST
- ☐ IBS_NASL
- ☐ GB

☐ Enforce inclusion

Tree Options

☒ Limit number of levels (tree depth)

Boosting Options

Number of models

Learning rate

OK Apply Cancel ?

Dialog - 3:21 - Gradient Boosted Trees Predictor

File

Prediction Settings Flow Variables Job Manager Selection Memory Policy

☐ Change prediction column name

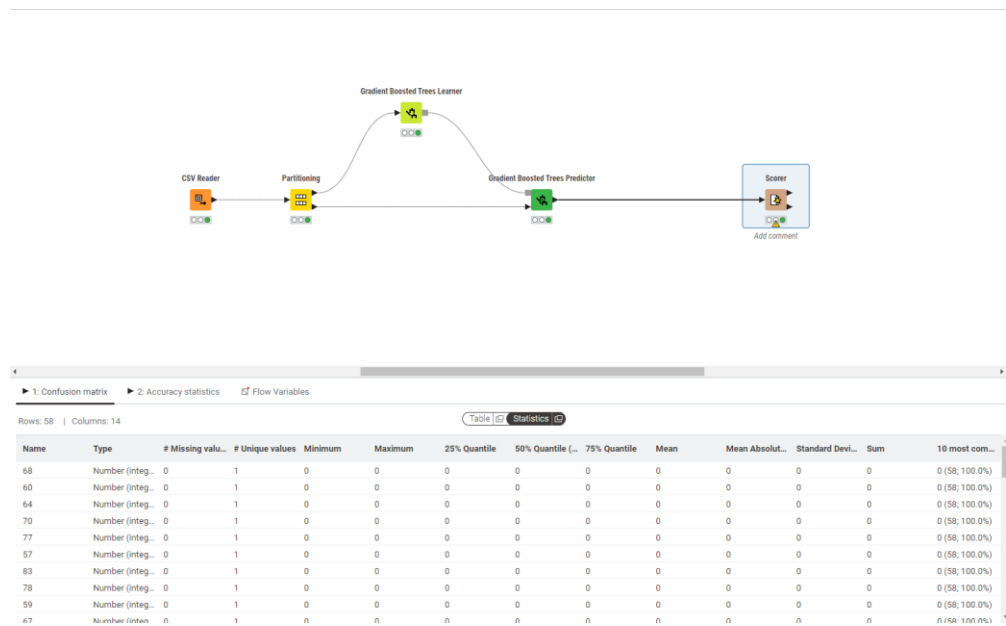
Prediction column name

☒ Append overall prediction confidence

☐ Append individual class probabilities

Suffix for probability columns

OK Apply Cancel ?



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.