

SPRAWOZDANIE

Zajęcia: Uczenie Maszynowe

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 1 Data 05.10.2024 Temat: "Praktyczne zastosowanie regresji liniowej w analizie danych. Implementacja algorytmów klasyfikacji binarnej w Pythonie" Wariant 11	Szymon Nycz Informatyka II stopień, niestacjonarne, 1 semestr, gr.1b
---	---

1. Polecenie:

Powikłania zawału mięśnia sercowego:

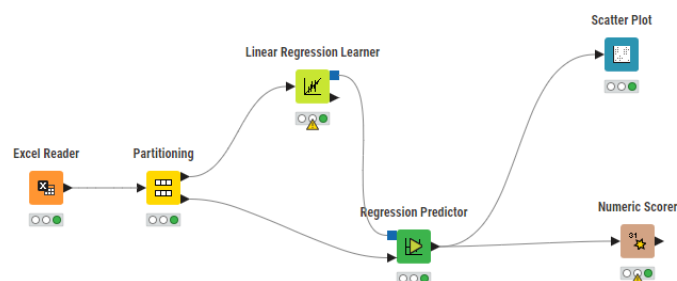
<https://www.kaggle.com/datasets/rafatashrafjoy/myocardial-infarction-complications>

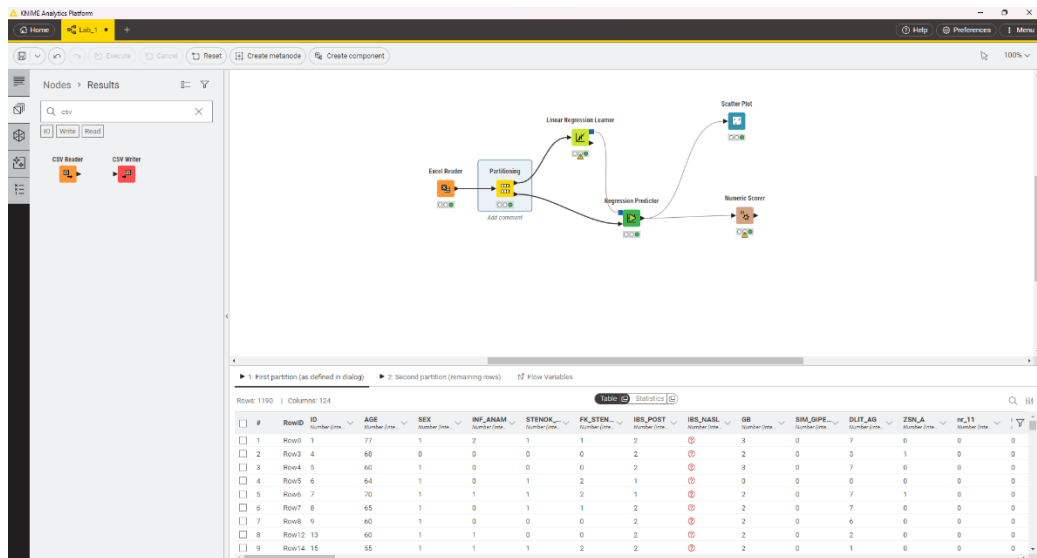
2. Link do repozytorium:

Link: https://github.com/Maciek332/Semestr_1_Nycz/tree/master/UM/Lab_1

3. Opis programu opracowanego

Tworzenie przepływu pracy uczenia maszynowego w narzędziu KNIME





Dialog - 3:3 - Linear Regression Learner

File

Settings | Flow Variables | Job Manager Selection | Memory Policy

Target: IBS_NASL

Values

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

Filter

- ID
- AGE
- SEX
- INF_ANAM
- STENOK_AN
- FK_STENOK
- IBS_POST
- GB
- SIM_GPERT
- DLIT_AG
- ZSN_A

☐ Enforce inclusion

Regression Properties

☐ Predefined Offset Value: 0

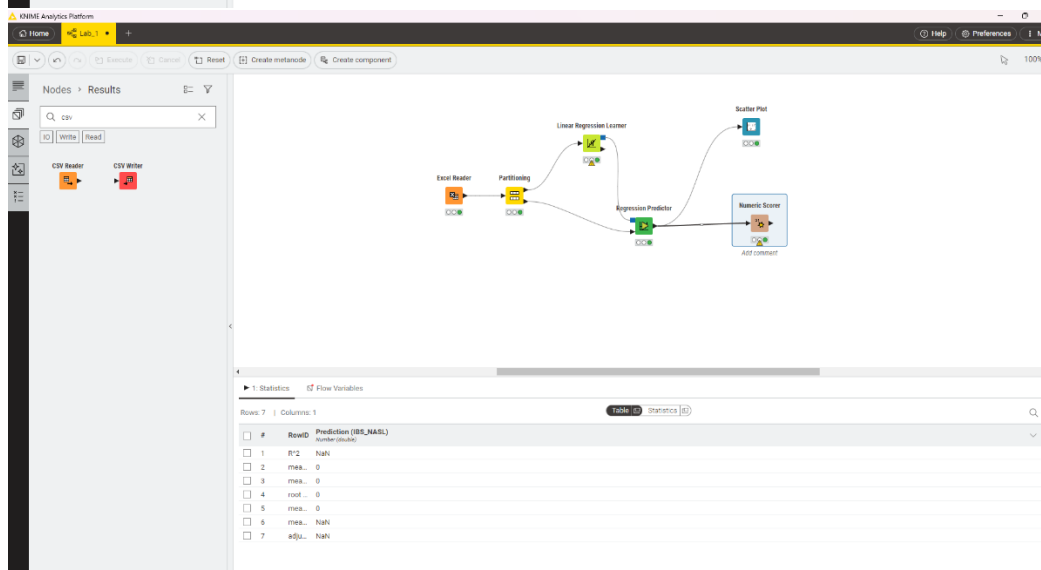
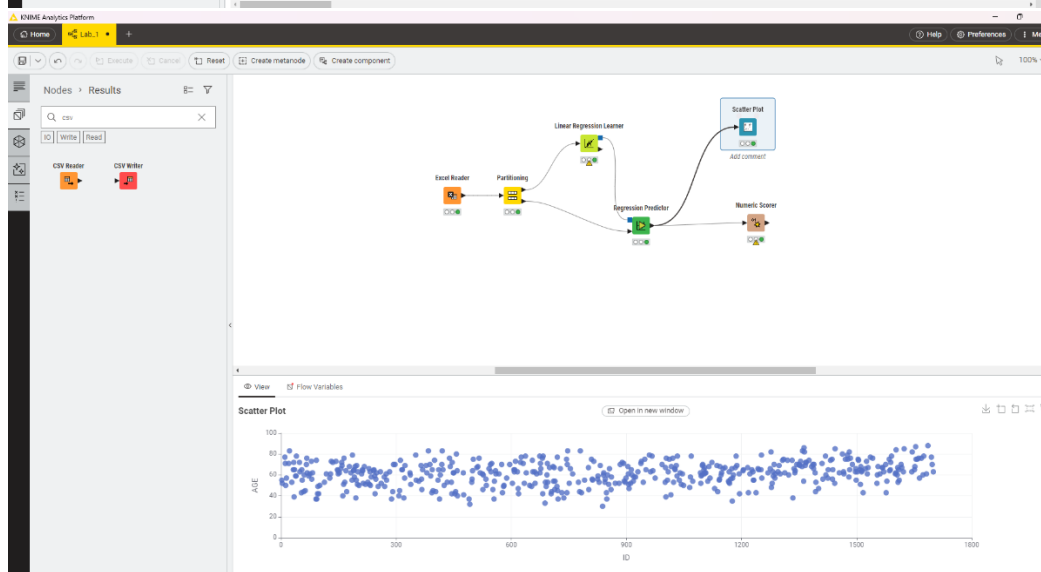
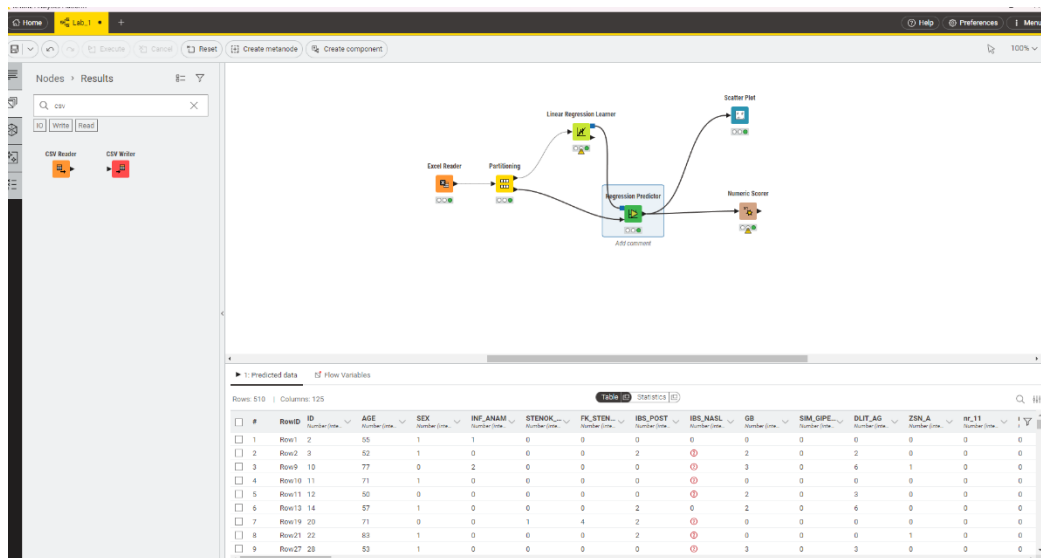
Missing Values in Input Data

☒ Ignore rows with missing values.
☐ Fail on observing missing values.

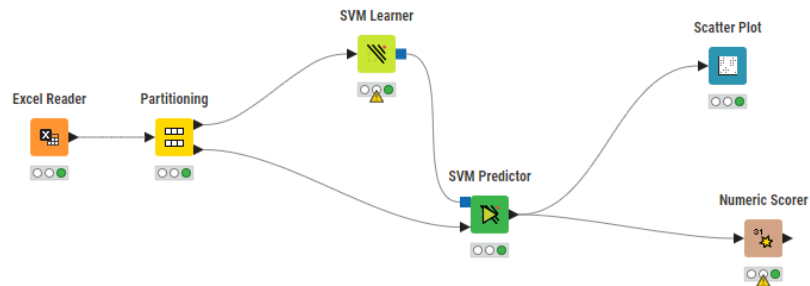
Scatter Plot View

First Row: 1
Row Count: 20 000

OK Apply Cancel ?



Opracowanie klasyfikacji binarnej



KNIME Analytics Platform

Nodes > Results

Search: num

Manipulation | Columns: +18

Number Rounder | Numeric Binzer | Numeric Scorer

Numeric Outliers | String to Number | Number to String

DE Numeric-Slicer | Numeric Distances | Category to Number

Duration to Number | Numeric Row Splitter | Number Format Manager

Edit Numeric Defaults | Numeric Binzer (PANEL) | Number to String (PANEL)

String to Number (PANEL) | Numeric Outliers (Apply) | Number to Category (Apply)

Category to Number (Apply) | Groupby

1. File Table | Flow Variables

Rows: 1700 | Columns: 124

#	RowID	ID	AGE	SEX	INF_ANAM	STENOK	FK_STEN	IBS_POST	IBS_NASL	GB	SIM_OIPE	DLIT_AG	ZSN_A	nr_11
1	Row0	1	77	1	2	1	1	2	0	3	0	7	0	0
2	Row1	2	55	1	1	0	0	0	0	0	0	0	0	0
3	Row2	3	52	1	0	0	0	2	0	2	0	2	0	0
4	Row3	4	68	0	0	0	0	2	0	2	0	3	1	0
5	Row4	5	60	1	0	0	0	2	0	3	0	7	0	0
6	Row5	6	64	1	0	1	2	1	0	0	0	0	0	0
7	Row6	7	70	1	1	1	2	1	0	2	0	7	1	0
8	Row7	8	65	1	0	1	1	2	0	2	0	7	0	0
9	Row8	9	60	1	0	0	0	2	0	2	0	6	0	0

KNIME Analytics Platform

Nodes > Results

Search: num

Manipulation | Columns: +18

Number Rounder | Numeric Binzer | Numeric Scorer

Numeric Outliers | String to Number | Number to String

DE Numeric-Slicer | Numeric Distances | Category to Number

Duration to Number | Numeric Row Splitter | Number Format Manager

Edit Numeric Defaults | Numeric Binzer (PANEL) | Number to String (PANEL)

String to Number (PANEL) | Numeric Outliers (Apply) | Number to Category (Apply)

Category to Number (Apply) | Groupby

1. First partition (as defined in dialog) | 2. Second partition (remaining rows) | Flow Variables

Rows: 1190 | Columns: 124

#	RowID	ID	AGE	SEX	INF_ANAM	STENOK	FK_STEN	IBS_POST	IBS_NASL	GB	SIM_OIPE	DLIT_AG	ZSN_A	nr_11
1	Row0	1	77	1	2	1	1	2	0	3	0	7	0	0
2	Row2	3	52	1	0	0	0	2	0	2	0	2	0	0
3	Row3	4	68	0	0	0	0	2	0	2	0	3	1	0
4	Row5	6	64	1	0	1	2	1	0	0	0	0	0	0
5	Row6	7	70	1	1	1	2	1	0	2	0	7	1	0
6	Row7	8	65	1	0	1	1	2	0	2	0	7	0	0
7	Row9	10	60	1	0	0	0	2	0	2	0	6	0	0
8	Row10	11	71	1	0	0	0	0	0	0	0	0	0	0
9	Row11	12	50	0	0	0	0	0	0	2	0	3	0	0

Dialog - 3:9 - SVM Learner

File

Options | Flow Variables | Job Manager Selection

Class column:

Overlapping penalty:

Choose your kernel and parameters:

☒ Polynomial

Power: Bias: Gamma:

☐ HyperTangent

kappa: delta:

☐ RBF

sigma:

OK Apply Cancel ?

KNIME Analytics Platform

Nodes > Results

Search: num

Manipulation | Columns | v18

Nodes: Number Reader, Numeric Reader, Numeric Scorer, String to Number, Number to String, DB Numeric Reader, Numeric Distance, Category to Number, Duration to Number, Numeric Row Scorer, Number Format Manager, Add Numeric Details, Numeric Scorer (PANEL), Number to String (PANEL), String to Number (PANEL), Numeric Outliers (PANEL), Number to Category (PANEL), Category to Number (PANEL), GroupBy

Workflow:

```

graph LR
    LR[Local Reader] --> P[Porting]
    P --> SL[SVM Learner]
    P --> SP[SVM Predictor]
    SP --> SPLOT[Scatter Plot]
    SP --> NS[Numeric Scorer]
  
```

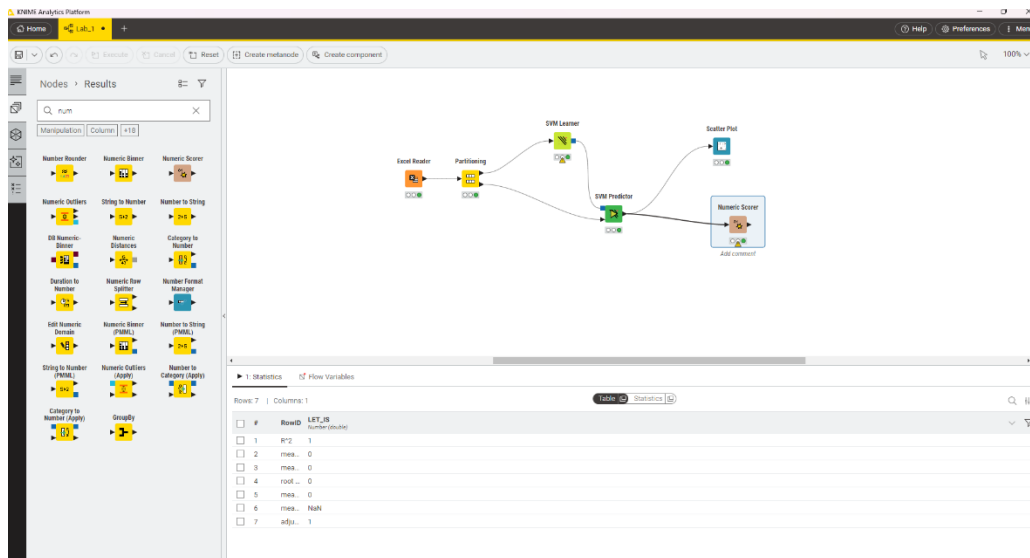
Classification

Flow Variables

Rows: 510 | Columns: 125

#	RowID	AGE	SEX	IMP_ANAM	STENOC	FK_STEN	IES_POST	IES_MAIL	GB	SIM_SINE	DLT_AD	ZIN_A	HY_11
1	Row1 2	55	1	1	0	0	0	0	0	0	0	0	0
2	Row4 5	60	1	0	0	0	2	0	0	0	7	0	0
3	Row9 10	77	0	2	0	0	0	0	3	0	5	1	0
4	Row12 13	60	1	1	0	0	2	0	2	0	2	0	0
5	Row14 15	55	1	1	1	2	2	0	2	0	1	0	0
6	Row19 20	71	0	0	1	4	2	0	0	0	0	0	0
7	Row30 31	64	0	0	0	0	2	0	2	0	1	0	0
8	Row32 33	72	0	0	0	0	2	0	2	0	7	0	0
9	Row35 37	44	0	0	5	2	1	0	3	0	2	1	0





4. Wnioski

Przepływ pracy w KNIME opiera się na łączeniu węzłów, które odzwierciedlają przepływ danych. Proces rozpoczyna się od załadowania danych, które następnie są przetwarzane, jeśli wymaga tego dalsza analiza. Przed budową modelu uczenia maszynowego dane często dzielone są na zestawy treningowe i testowe, aby umożliwić ocenę jakości modelu. KNIME oferuje szeroki wybór algorytmów uczenia maszynowego, takich jak regresja liniowa, drzewa decyzyjne, lasy losowe i maszyny wektorów nośnych. Po stworzeniu modelu jego skuteczność jest oceniana na podstawie zestawu testowego, a bardziej rzetelna ocena może być osiągnięta dzięki walidacji krzyżowej.

KNIME to wszechstronne narzędzie do analizy danych i modelowania uczenia maszynowego, które umożliwia szybkie tworzenie i ocenę modeli. Dzięki intuicyjnemu interfejsowi graficznemu oraz rozbudowanej bazie węzłów użytkownicy mogą tworzyć złożone przepływy pracy bez potrzeby programowania.