

# 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 10

Anna Więzik

Informatyka

II stopień, niestacjonarne,  
1 semestr, gr.1b

## 1. Polecenie:

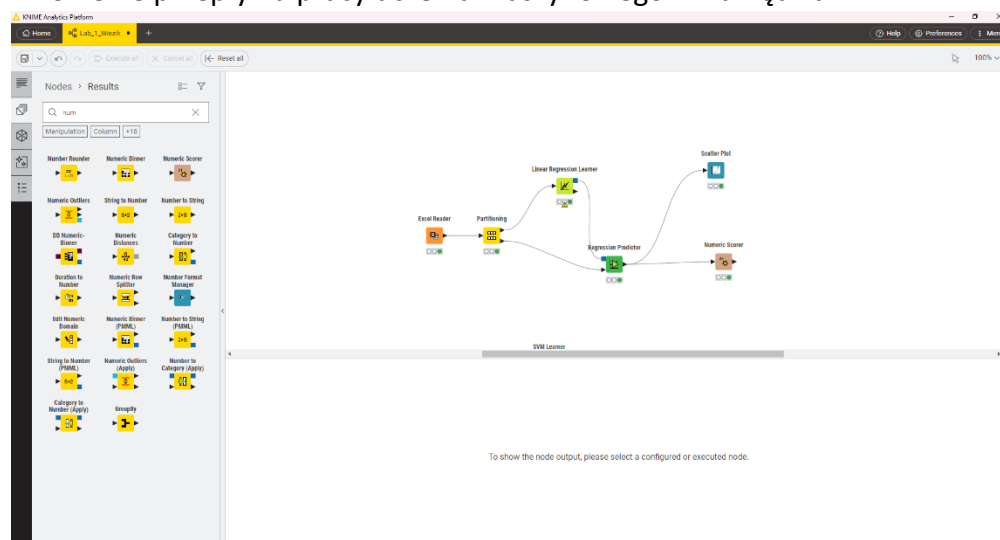
Smoking patient: <https://www.kaggle.com/datasets/thomaskonstantin/cpg-values-of-smoking-and-non-smoking-patients>

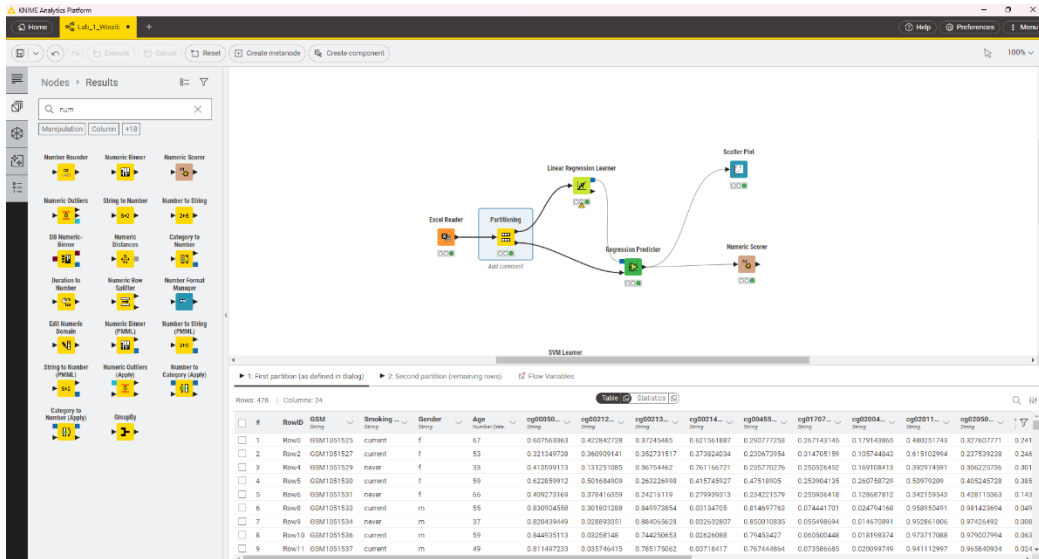
## 2. Link do repozytorium:

Link: [https://github.com/AnaShiro/UM\\_2024](https://github.com/AnaShiro/UM_2024)

## 3. Opis programu opracowanego

Tworzenie przepływu pracy uczenia maszynowego w narzędziu KNIME





Dialog - 3:3 - Linear Regression Learner

Settings | Flow Variables | Job Manager Selection | Memory Policy

Target: Age

Values: Manual Selection (selected) | Wildcard/Regex Selection

Exclude: Filter (No columns in this list)

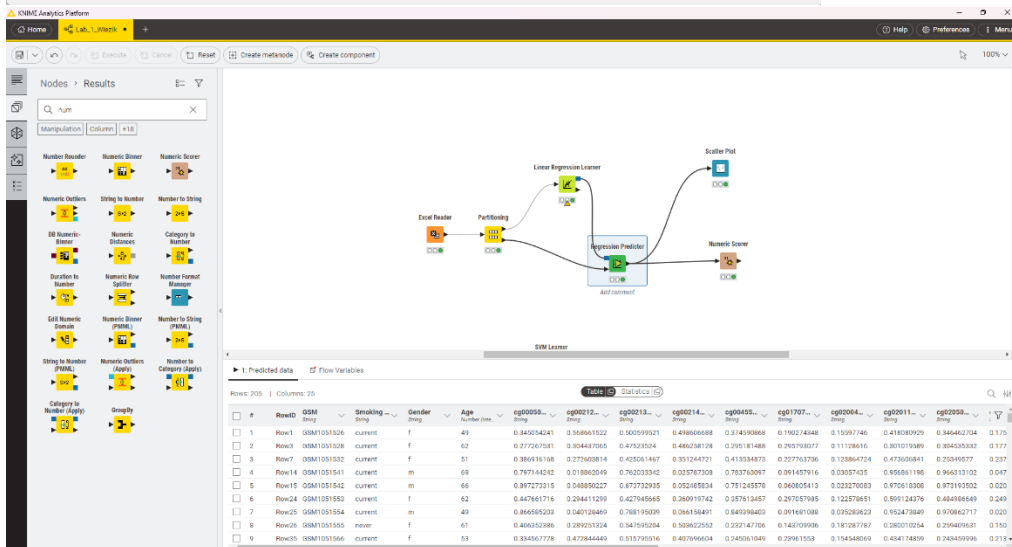
Include: GSM, Smoking Status, Gender, cg00050873, cg00212031, cg00213748, cg00214611, cg00455876, cg01707559, cg02004872

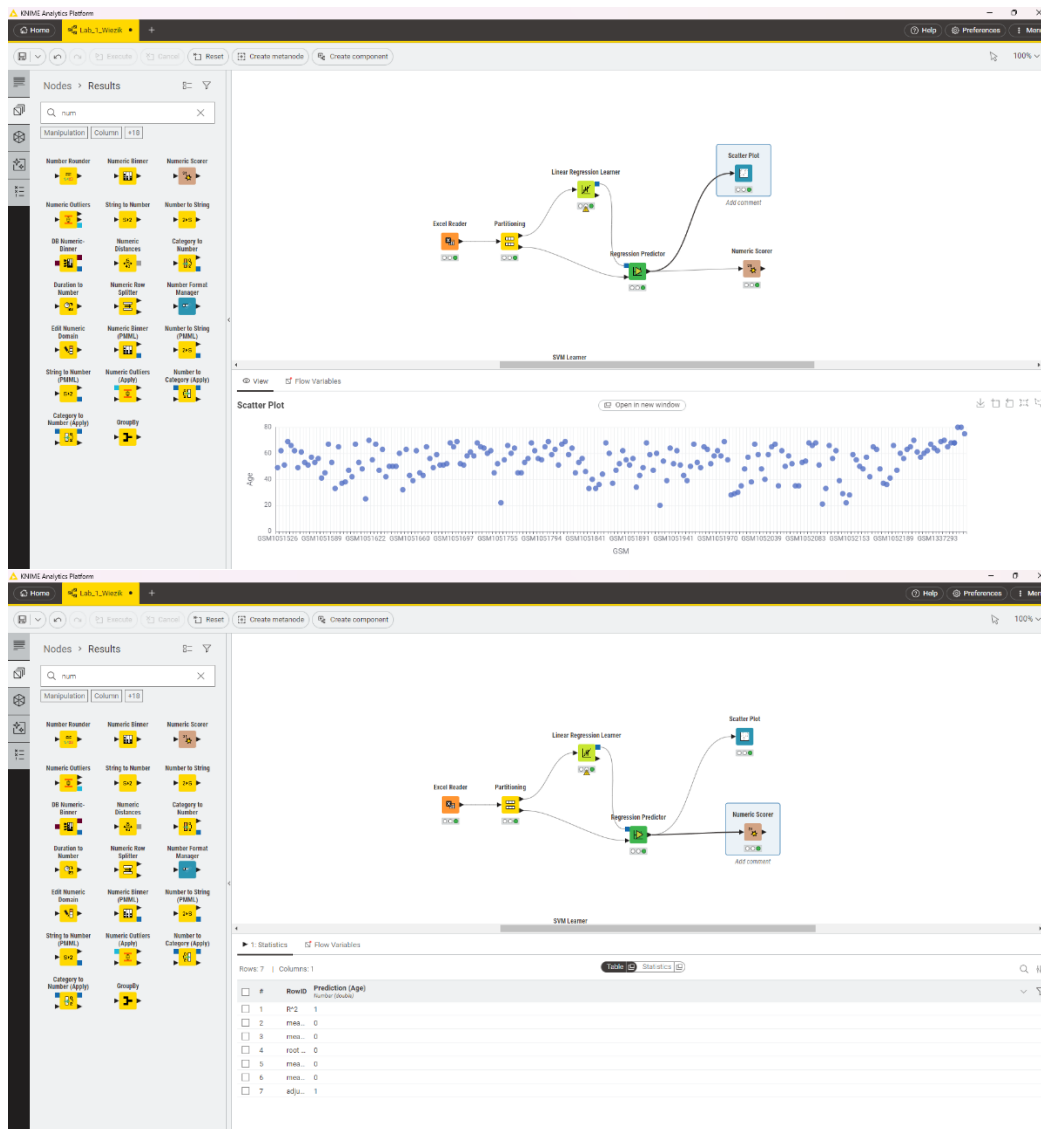
Regression Properties: Predefined Offset Value: 0

Missing Values in Input Data: Ignore rows with missing values. (selected)

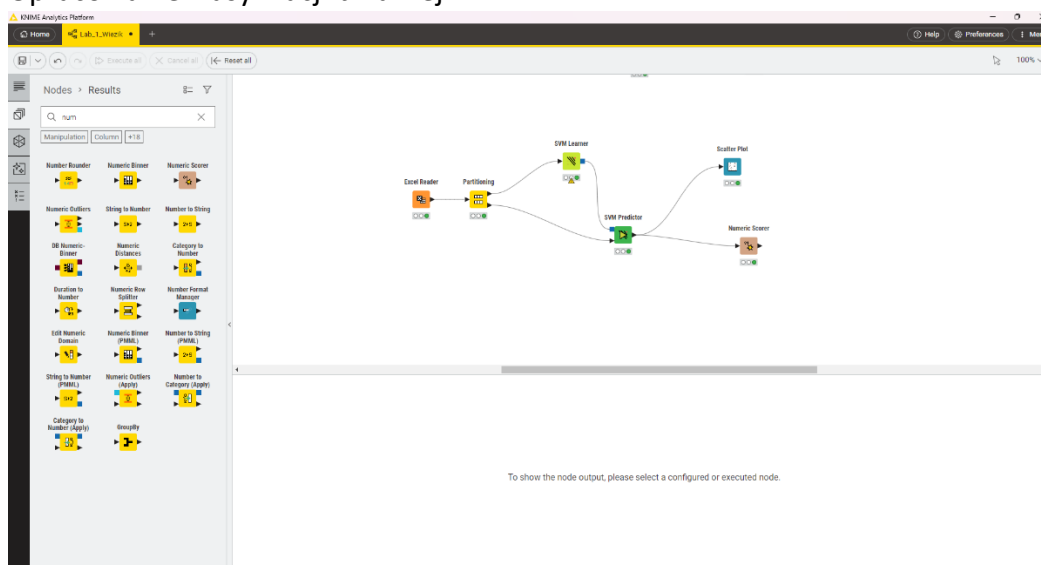
Scatter Plot View: First Row: 1, Row Count: 20 000

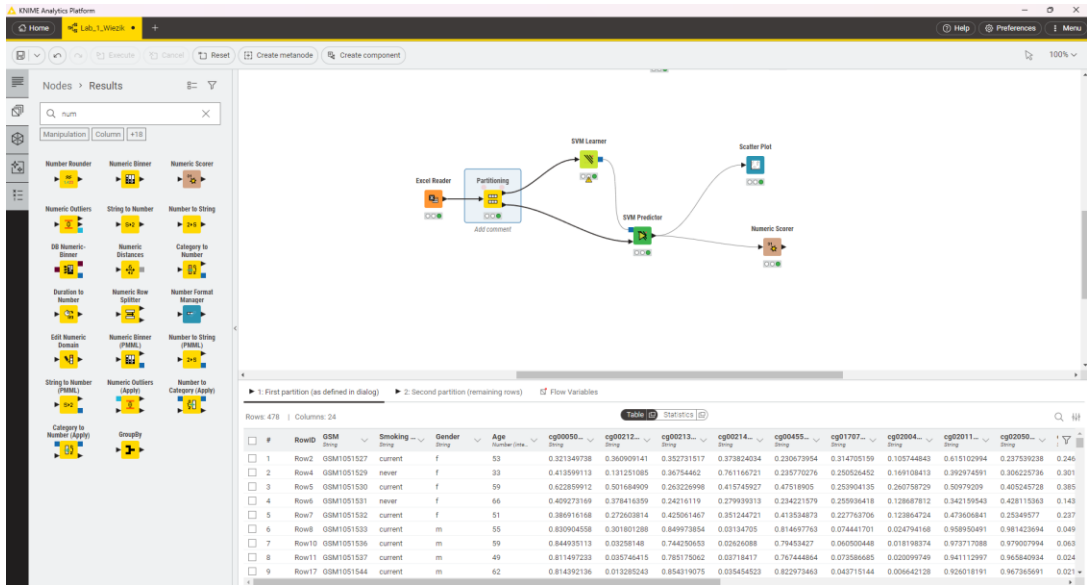
OK Apply Cancel ?





## Opracowanie klasyfikacji binarnej





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:

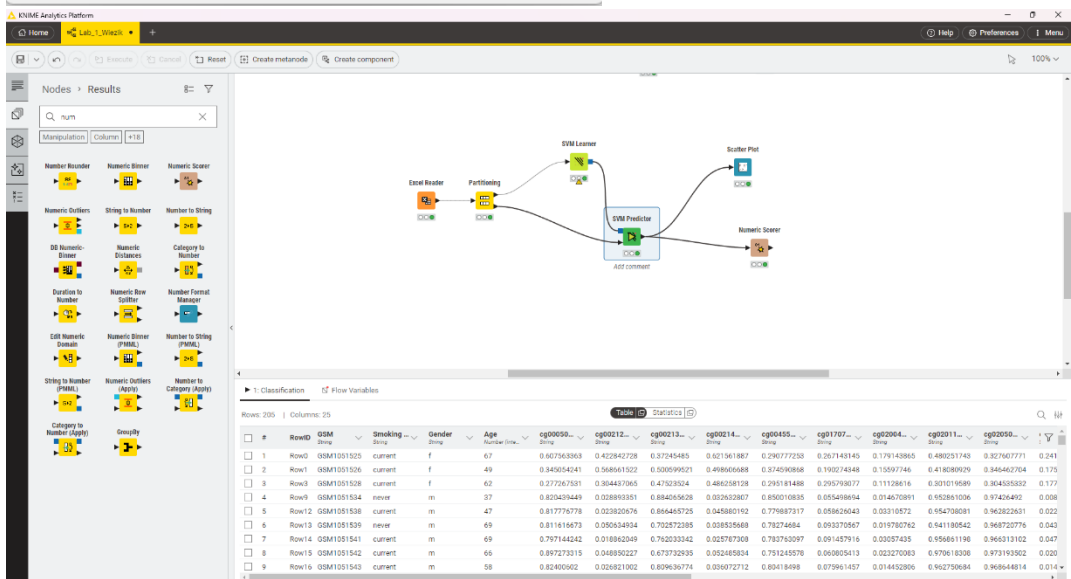
☐ HyperTangent

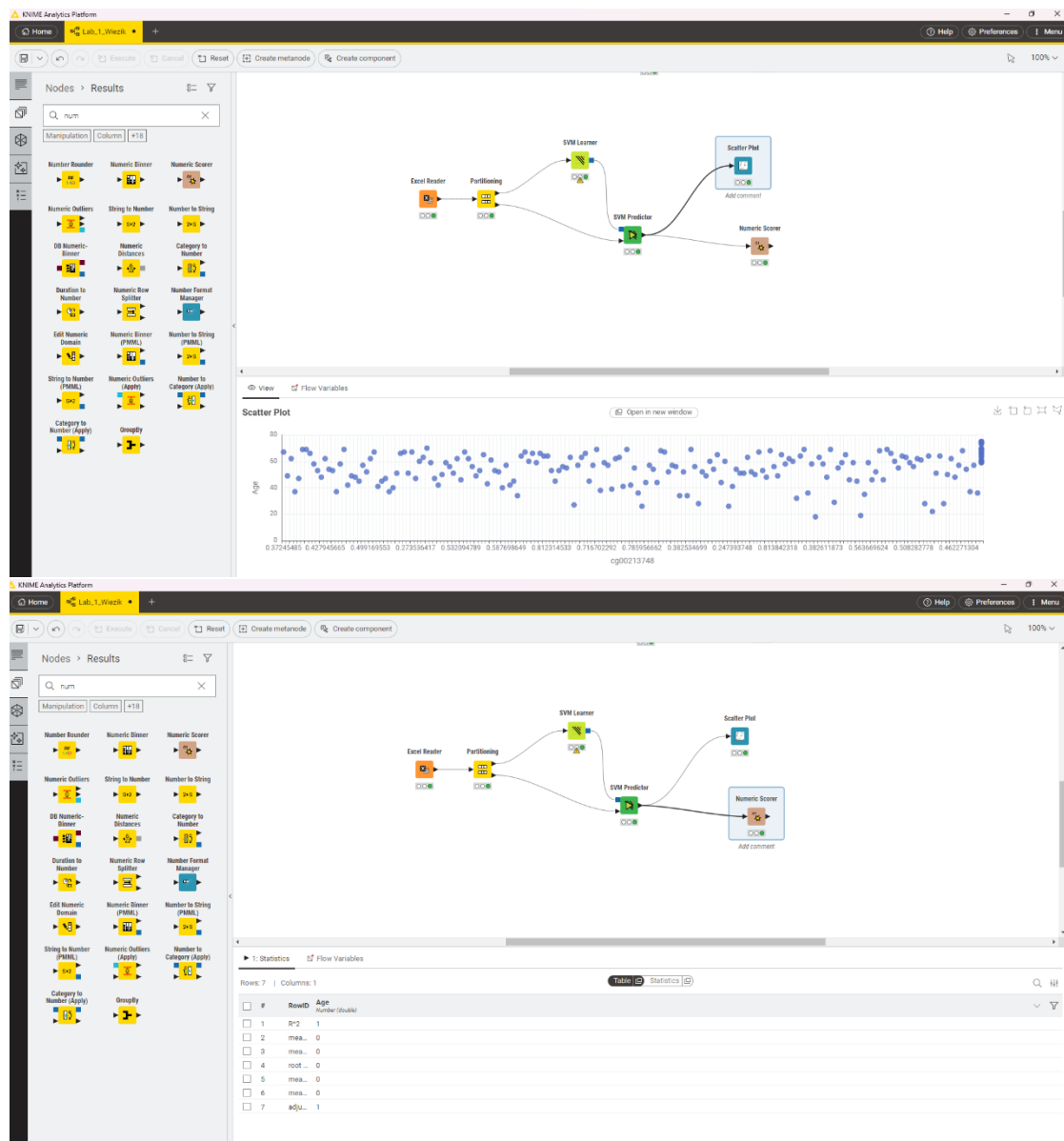
kappa:  delta:

☐ RBF

sigma:

OK Apply Cancel ?





## 4. Wnioski

Przeptyw pracy w KNIME składa się z serii węzłów (ang. nodes), które są połączone w sposób reprezentujący przepływ danych. Pierwszym krokiem jest załadowanie danych, które będą analizowane i przetwarzane. Dane, które są zaimportowane, mogą wymagać przetworzenia przed zastosowaniem modelu uczenia maszynowego. W celu oceny jakości modelu uczenia maszynowego, dane są zazwyczaj dzielone na zbiór treningowy oraz testowy. KNIME wspiera szeroką gamę algorytmów uczenia maszynowego, takich jak regresja liniowa, drzewa decyzyjne, lasy losowe (ang. Random Forest), czy maszyny wektorów nośnych (ang. Support Vector Machines). Po zbudowaniu modelu, należy ocenić jego skuteczność na podstawie danych testowych. Aby uzyskać bardziej wiarygodną ocenę modelu, można zastosować walidację krzyżową (ang. cross-validation).

KNIME jest potężnym narzędziem do analizy danych i uczenia maszynowego, które pozwala na szybkie prototypowanie modeli i ich ocenę. Dzięki graficznemu interfejsowi oraz szerokiej bibliotece węzłów, KNIME umożliwia tworzenie zaawansowanych przepływów pracy bez konieczności programowania.