SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

|  |  |
| --- | --- |
| Laboratorium Nr 6  Data 07.12.2024  Temat: „Analiza danych z wykorzystaniem narzędzi do modelowania regresji”  Wariant 10 | Anna Więzik  Informatyka  II stopień, niestacjonarne,  1 semestr, gr.1b |

1. Polecenie:

Premise Child Health COVID-19 Health Services Disruption Survey 2020

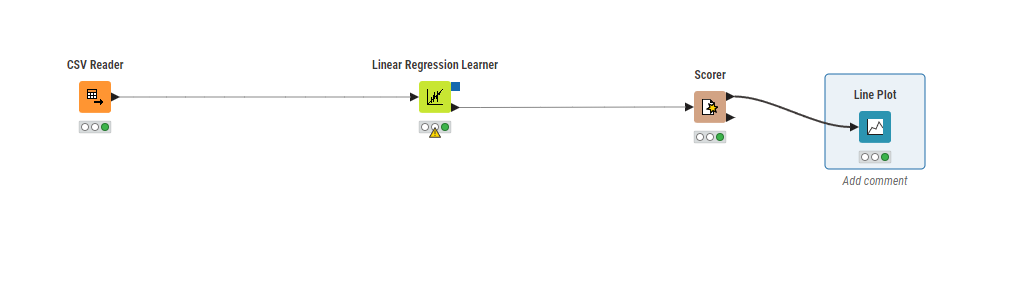
<http://ghdx.healthdata.org/record/ihme-data/premise-child-health-covid-19-health-services-disruption-survey-2020>

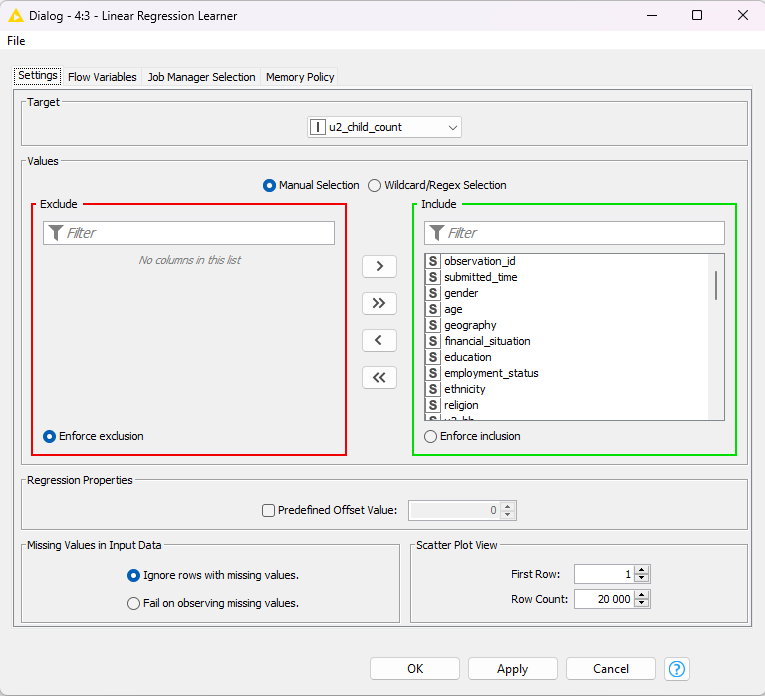
Link do repozytorium: <https://github.com/AnaShiro/NoD1_2024>

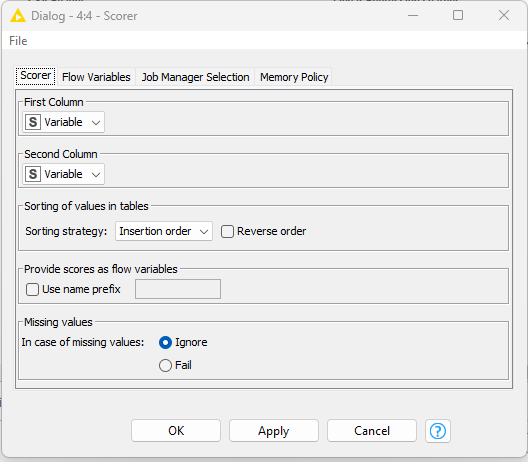
1. Opis programu opracowanego

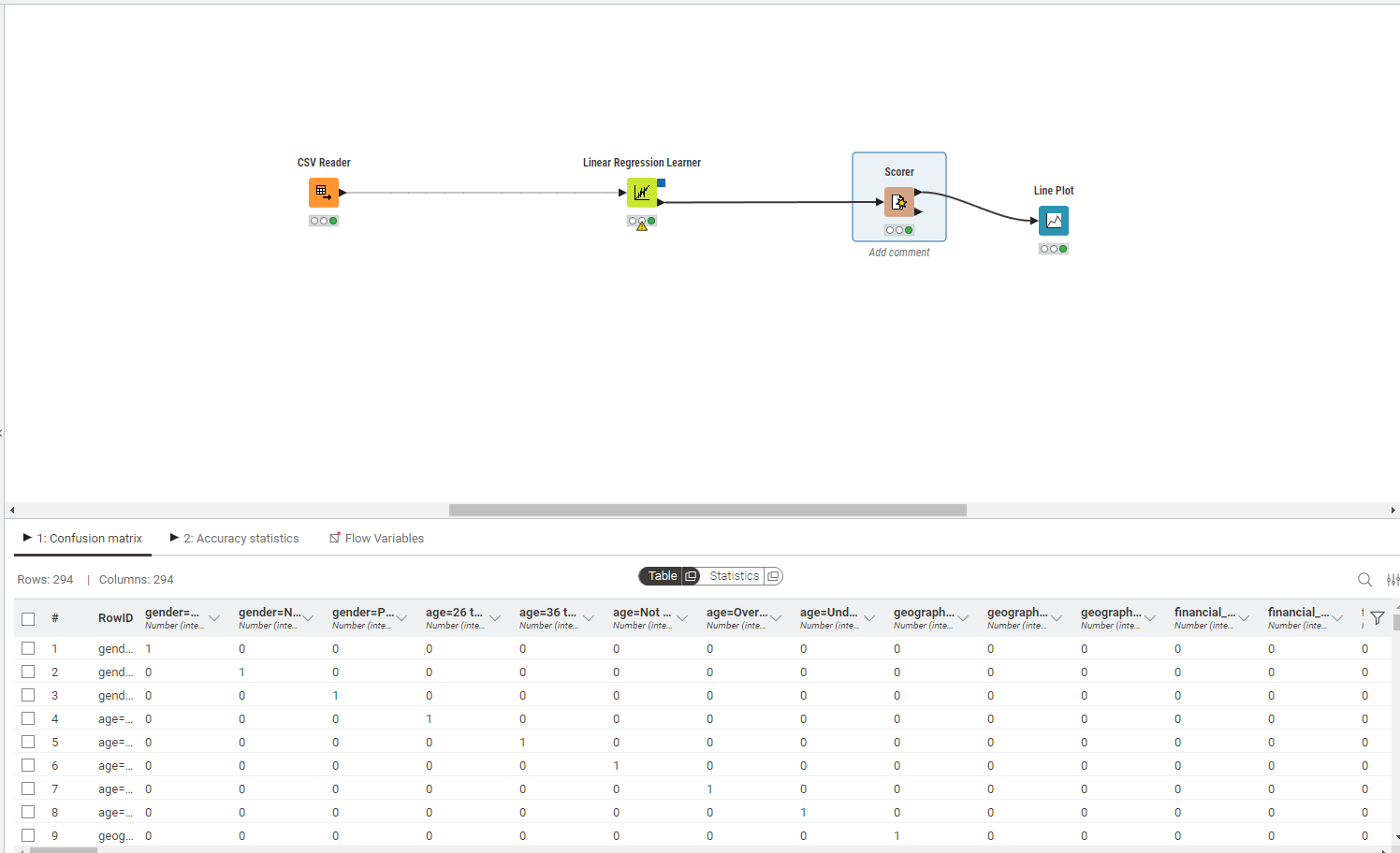
W Pythonie oraz KNIME porównaj wyniki regresji liniowej i Ridge na tym samym zbiorze danych.

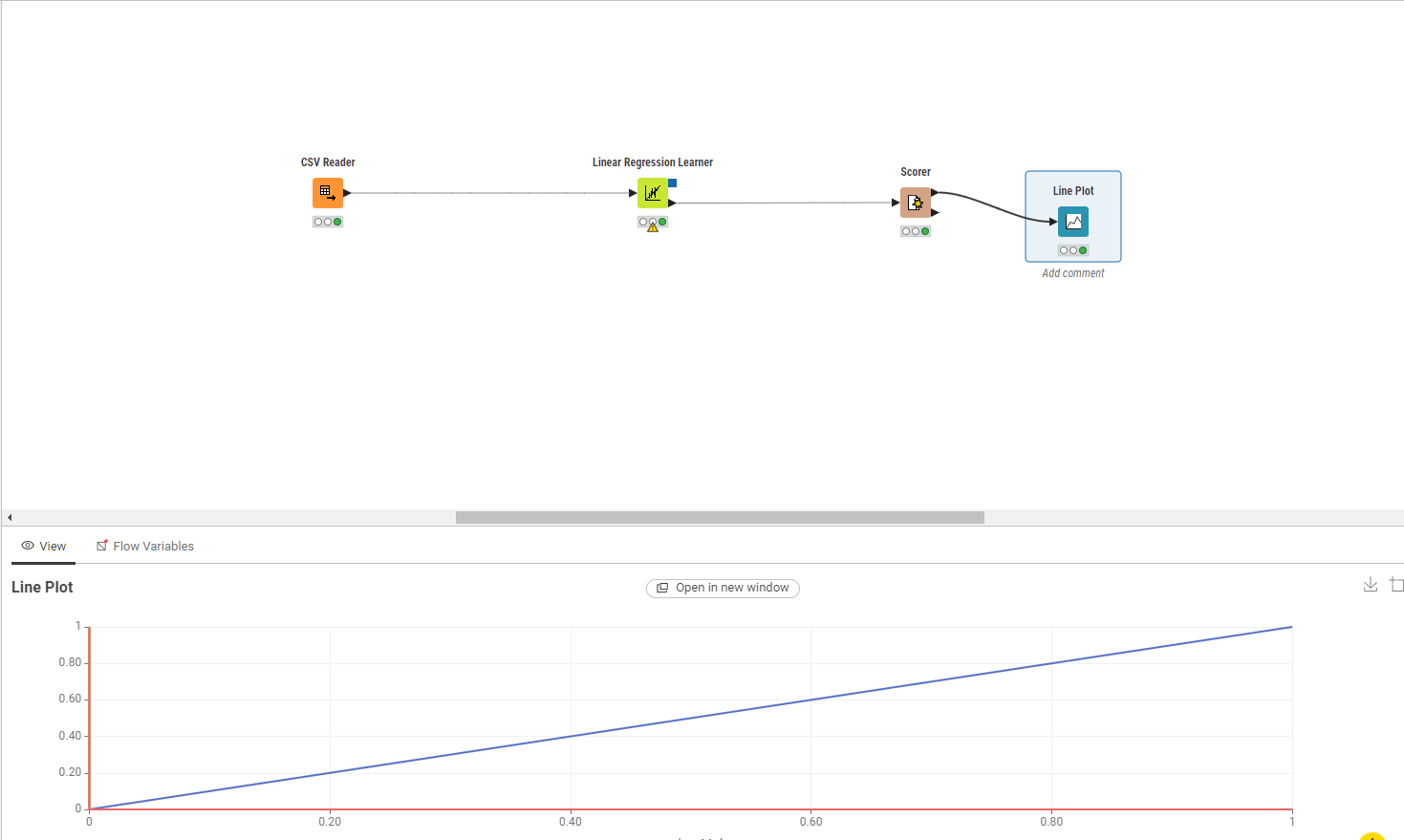
* KNIME



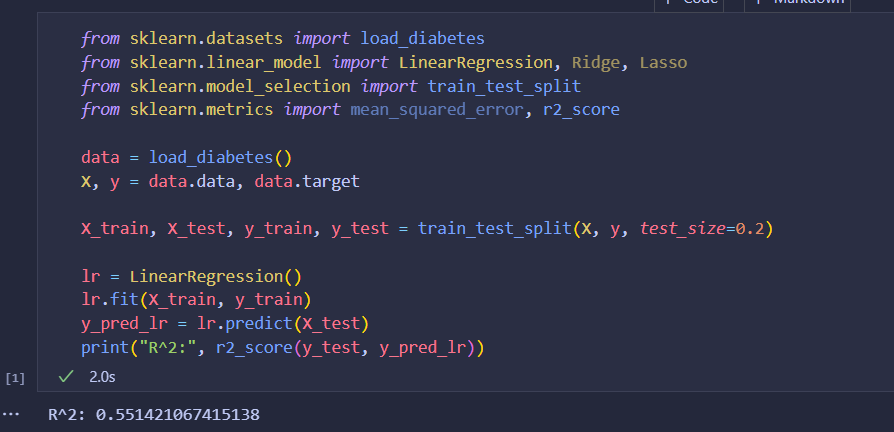






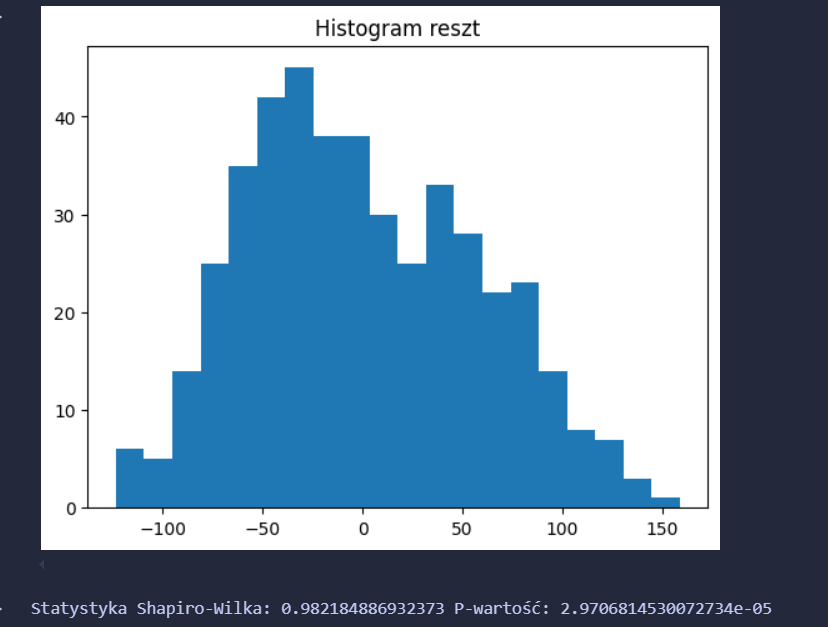


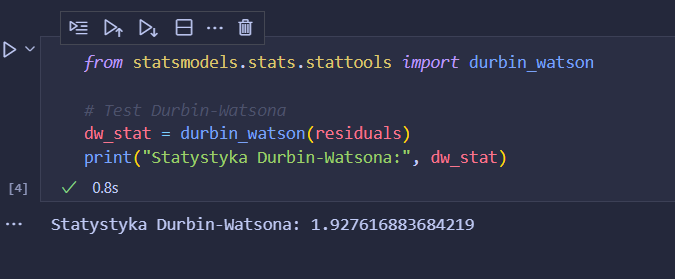
* Python



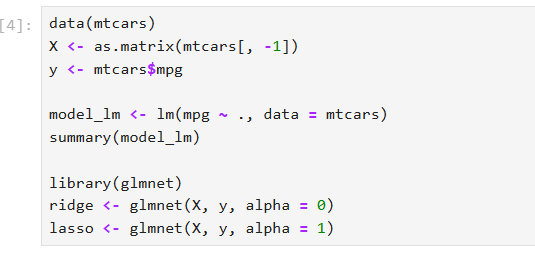


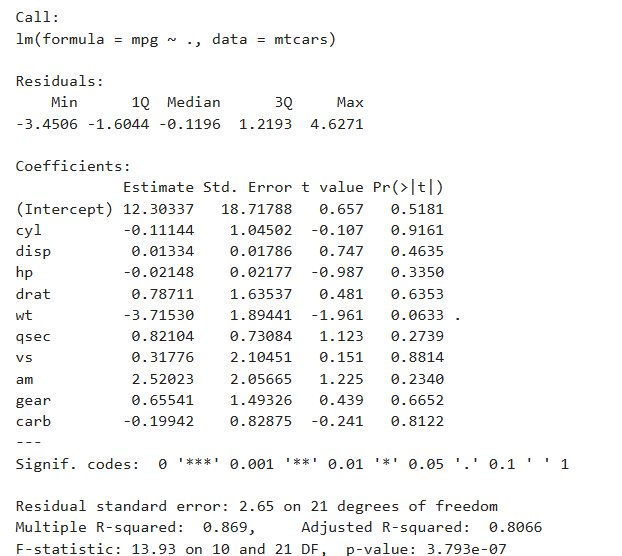


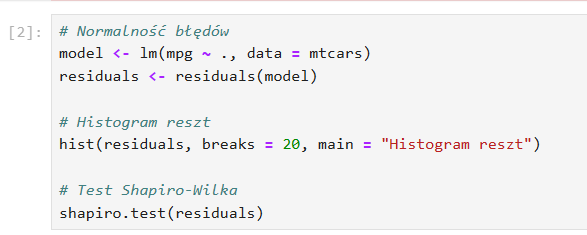


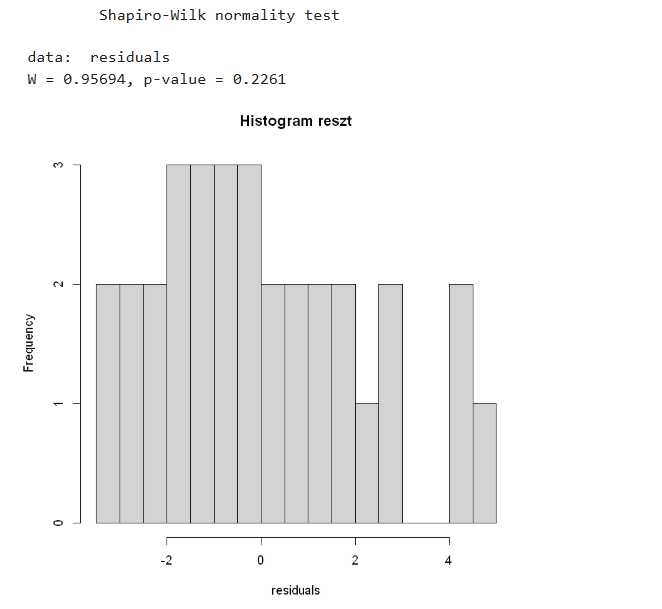


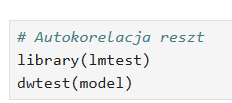
* R

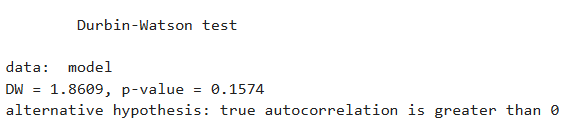












1. Wnioski

Regresja liniowa opisuje zależność zmiennej zależnej yy od zmiennych niezależnych x1,x2,…,xpx\_1, x\_2, \ldots, x\_p. W regresji grzbietowej (Ridge) współczynniki βj\beta\_j są regularizowane, aby podkreślić zmienne mające największy wpływ na wynik modelu. Reszty, czyli różnice między wartościami rzeczywistymi a przewidywanymi, oceniają jakość modelu. Regresja za pomocą sieci neuronowych wykorzystuje uczenie głębokie do modelowania złożonych relacji między zmiennymi. W Pythonie można to osiągnąć za pomocą bibliotek TensorFlow i PyTorch. Wizualizacja sieci neuronowych, na przykład przy użyciu plot\_model w TensorFlow, jest kluczowa dla zrozumienia architektury modelu. Rozmiar partii (batch size) określa liczbę próbek przetwarzanych jednocześnie w jednej aktualizacji wag, epoka to jedno pełne przejście przez zestaw danych treningowych, a iteracja to pojedyncza aktualizacja wag.