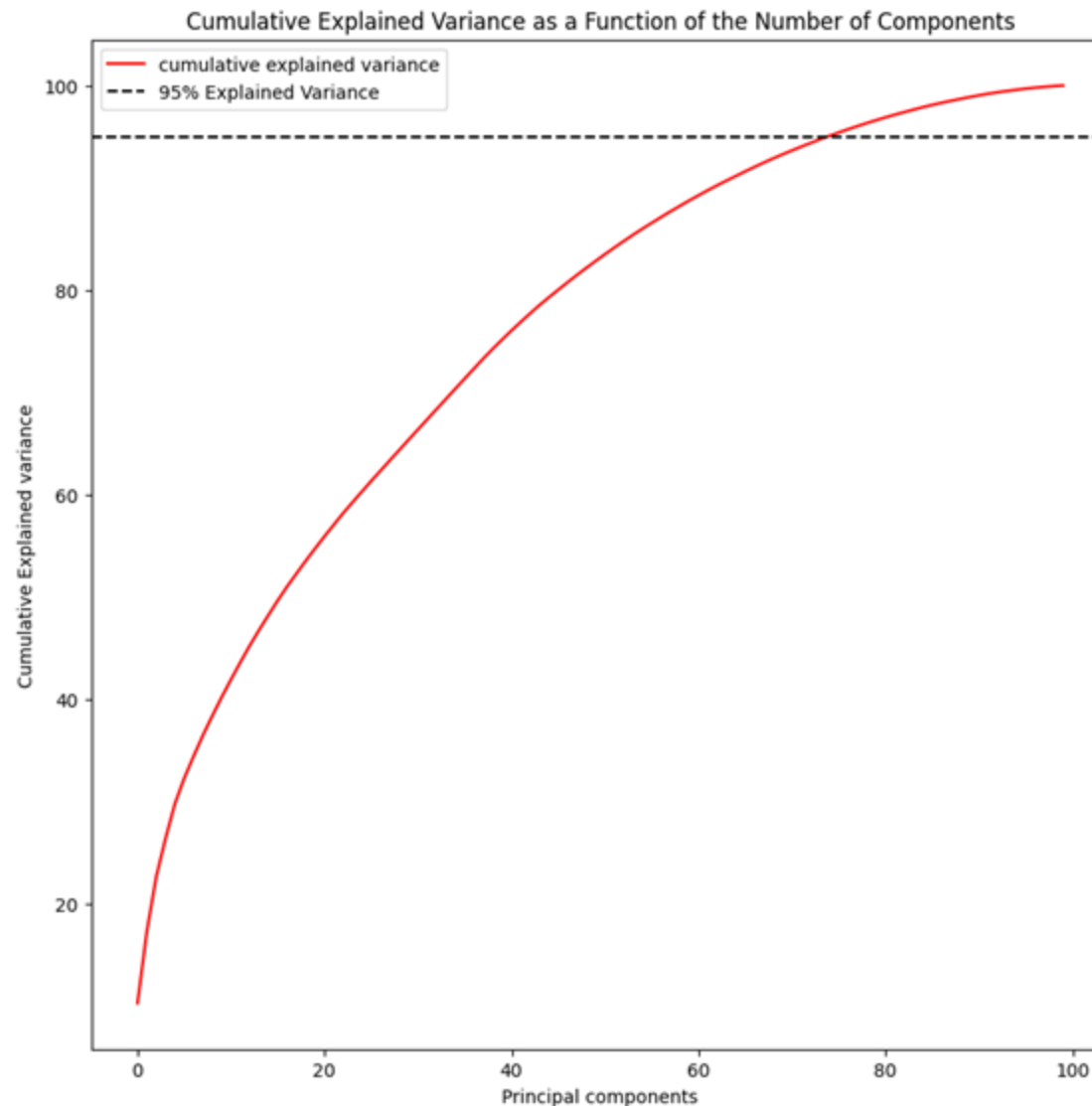

Uczenie maszynowe na wesoło

Czyli jak dopasować duuuuże dane dla problemu regresji?

Patrycja Kielan
Mateusz Zacharecki

Eksploracja danych

- Brak braków danych
- Zmienne typu float
- Redukcja wymiaru metodą PCA nie daje oczekiwanych rezultatów



Metody selekcji zmiennych

- Wybór zmiennych z wykorzystaniem modelu drzewa regresyjnego
 - Wrapper (10 zmiennych)
 - Filter (10 zmiennych + wykorzystanie drzewa)
 - Wrapper + filter (wybór 20 zmiennych za pomocą wrappera, następnie wybór 5 zmiennych za pomocą filtra)
-

Wykorzystane modele

- Gradient Boosting (wrapper)
 - SVM
 - Sieć neuronowa
 - XGBoost
 - Regresja liniowa
 - Voting (Select From Model, Wrapper, Wrapper + Filter)
-

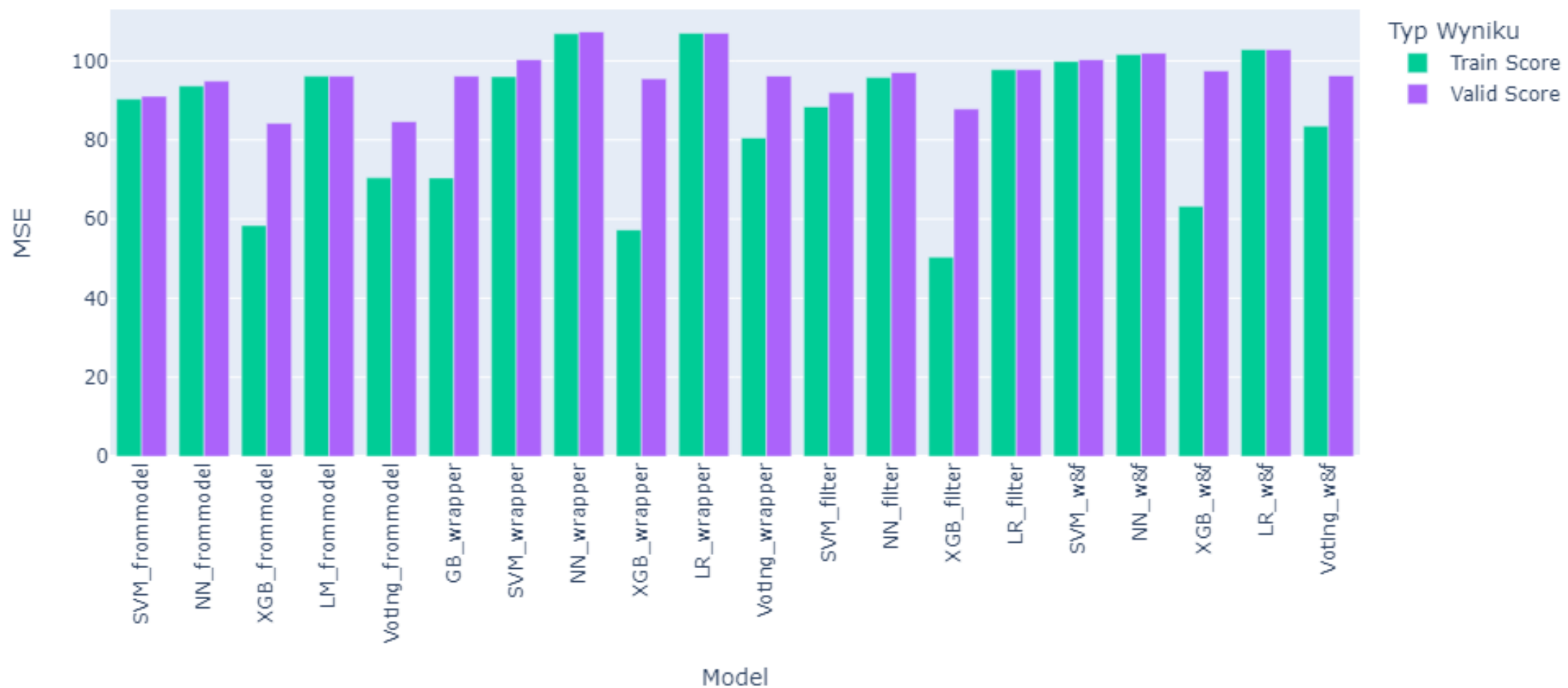
Inne podejścia

- Wykorzystanie trzykrotnej krosvalidacji
- Dobór parametrów przy użyciu siatki parametrów (Gradient Boosting, sieć neuronowa, XGBoost)

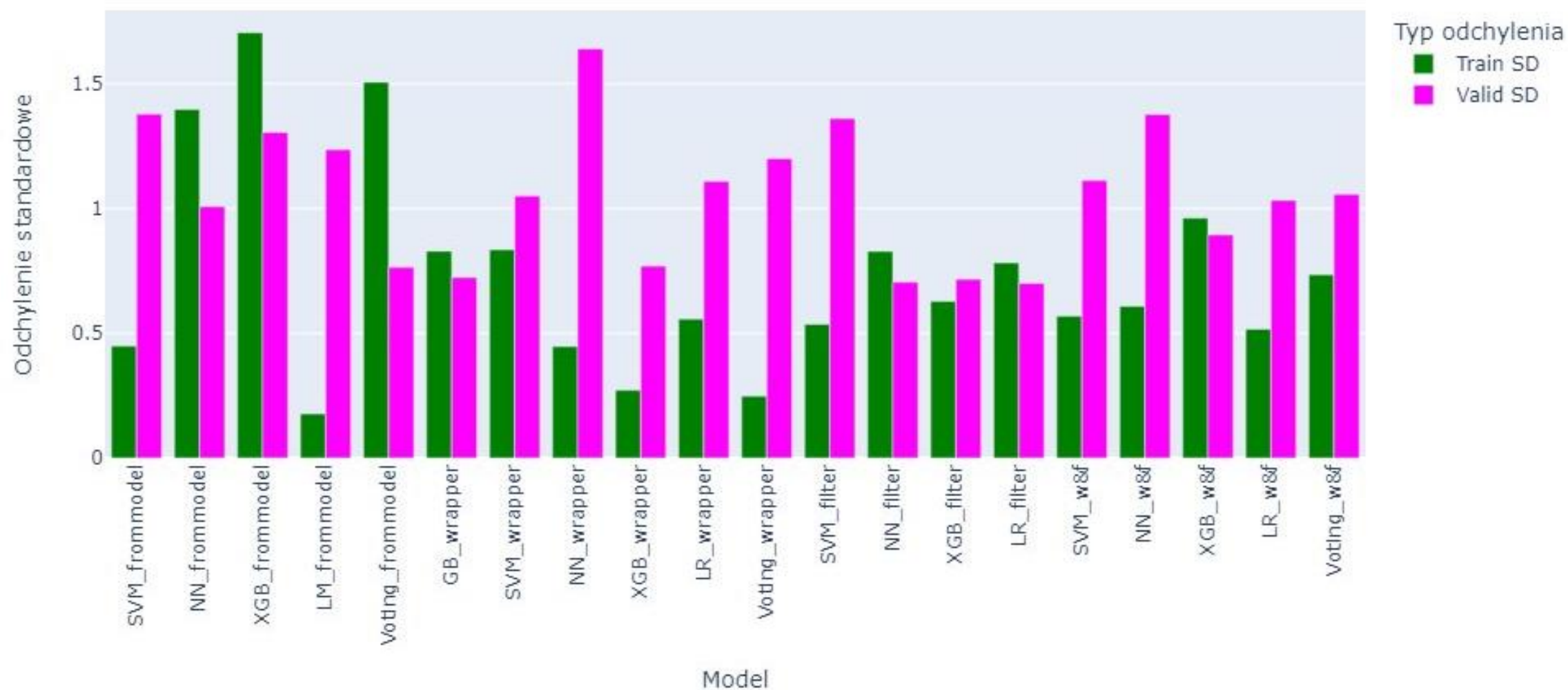


Wyniki

MSE scores for selected models



Odchylenia standardowe MSE dla wybranych modeli



Dziękujemy za uwagę!