

Raport

Wynik

Proponowanym rozwiązaniem jest klasyfikator xgBoost z odpowiednio wyznaczonymi parametrami. Przy pomocy miary zdefiniowanej w zadaniu osiągnął on średni wynik 84.71% dla dziesięciokrotnej krosvalidacji.

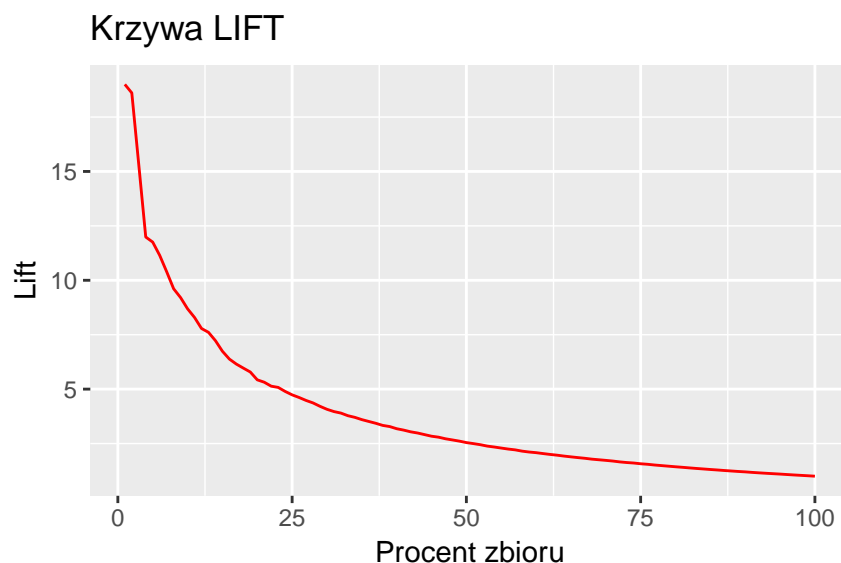
Lista rozważanych modeli/zmiennych

W pracy rozważono również inne modele. Dla każdego z nich wybrano odpowiednie zmienne za pomocą algorytmów charakterystycznych dla modelu (m. in. metryki ważności zmiennych). Poniżej przedstawiona jest tabela z właściwościami analizowanych modeli.

Model	Wynik	Liczba wybranych zmiennych
Regresja logistyczna	72.92%	8
Naiwny klasyfikator Bayesa	76.92%	12
Las losowy	82.21%	32
Reguły C5.0	84.08%	16
xgBoost	84.75%	9

Wykres

Poniżej przedstawiamy zachowanie rozwiązania dla danych treningowych.



Argumentacja poprawności

Do mierzenia jakości modeli użyto dziesięciokrotnej krosvalidacji zbioru treningowego.