

# Podsumowanie danych

## Surowy dataset

	rownames	gender	ethnicity	score	...	tuition	education	income	region
0	1	male	other	39.150002	...	0.88915	12	high	other
1	2	female	other	48.869999	...	0.88915	12	low	other
2	3	male	other	48.740002	...	0.88915	12	low	other
3	4	male	afam	40.400002	...	0.88915	12	low	other
4	5	female	other	40.480000	...	0.88915	13	low	other
...	...	...	...	...	...	...	...	...	...
4734	9391	male	afam	56.529999	...	0.25751	13	high	west
4735	9401	male	afam	59.770000	...	0.25751	15	high	west
4736	9411	male	other	43.169998	...	0.25751	12	high	west
4737	9421	male	afam	49.970001	...	0.25751	16	high	west
4738	9431	male	afam	53.410000	...	0.25751	13	high	west

[4739 rows x 15 columns]

## Przygotowany dataset

	gender	score	...	ethnicity_hispanic	ethnicity_other
0	1	-0.869628	...	0	1
1	0	-0.167570	...	0	1
2	1	-0.176959	...	0	1
3	1	-0.779343	...	0	0
4	0	-0.773564	...	0	1
...	...	...	...	...	...
4734	1	0.385699	...	0	0
4735	1	0.619718	...	0	0
4736	1	-0.579271	...	0	1
4737	1	-0.088118	...	0	0
4738	1	0.160347	...	0	0

[4739 rows x 16 columns]

Usunięto 0 wierszy

Zmodyfikowano 0 wierszy

## Testowy dataset

gender	score	...	ethnicity_hispanic	ethnicity_other
--------	-------	-----	--------------------	-----------------

4416	0	-0.196461	...	0	1
1516	1	0.486096	...	0	1
4203	1	1.097147	...	0	1
4393	0	-0.239798	...	1	0
1422	0	-0.921632	...	0	0
...	...	...	...	...	...
1349	1	0.465872	...	0	1
2859	0	0.262911	...	0	1
915	0	-0.513543	...	0	1
4376	0	0.037559	...	1	0
3966	0	-0.287468	...	0	1

[1422 rows x 16 columns]

# Podsumowanie uczenia modelu

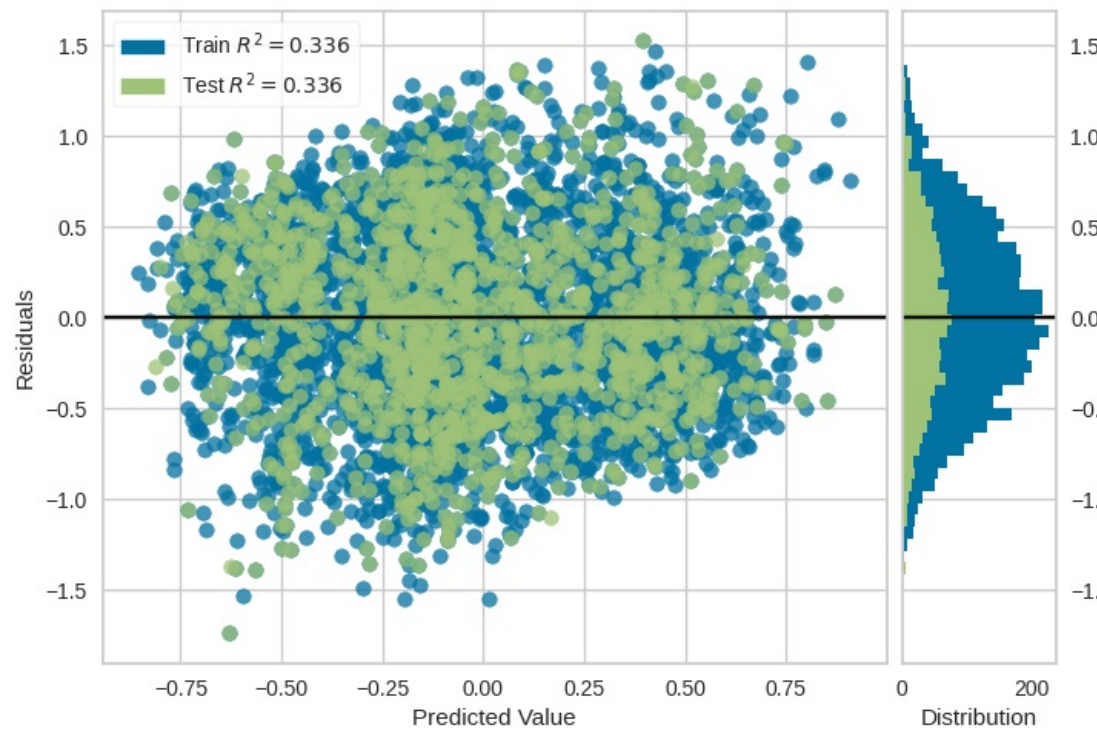
Wybrany model **Lars(random\_state=2024)**

Metryki modelu

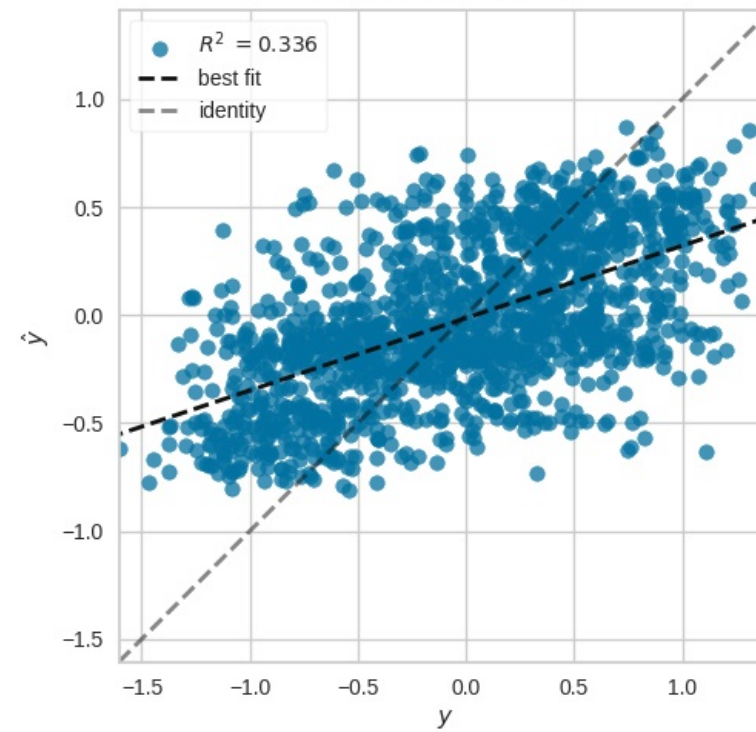
	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.4108	0.2599	0.5098	0.3355	0.2758	2.9068

Czas trenowania: 0:00:57.925591

Residuals for Lars Model



Prediction Error for Lars



Feature Importance Plot

