Prezentacja Projektu I Tunowalność

Katarzyna Mamla, Weronika Dyszkiewicz

Algorytmy

Algorytmy

W naszym projekcie rozważamy algorytmy:

- Random Forest,
- Light GBM,
- XgBoost.

Dla nich optymalizopwałyśmy parametry:

- Random Forest: *{'n_estimators', 'min_samples_split', 'min_samples_leaf', 'max_features','max_depth'}*
- Light GBM: *{num_leaves', 'n_estimators', learning_rate'}*,
- XgBoost: *{'subsample', 'min_child_weight', 'max_depth', 'eta','colsample_bytree'}*.

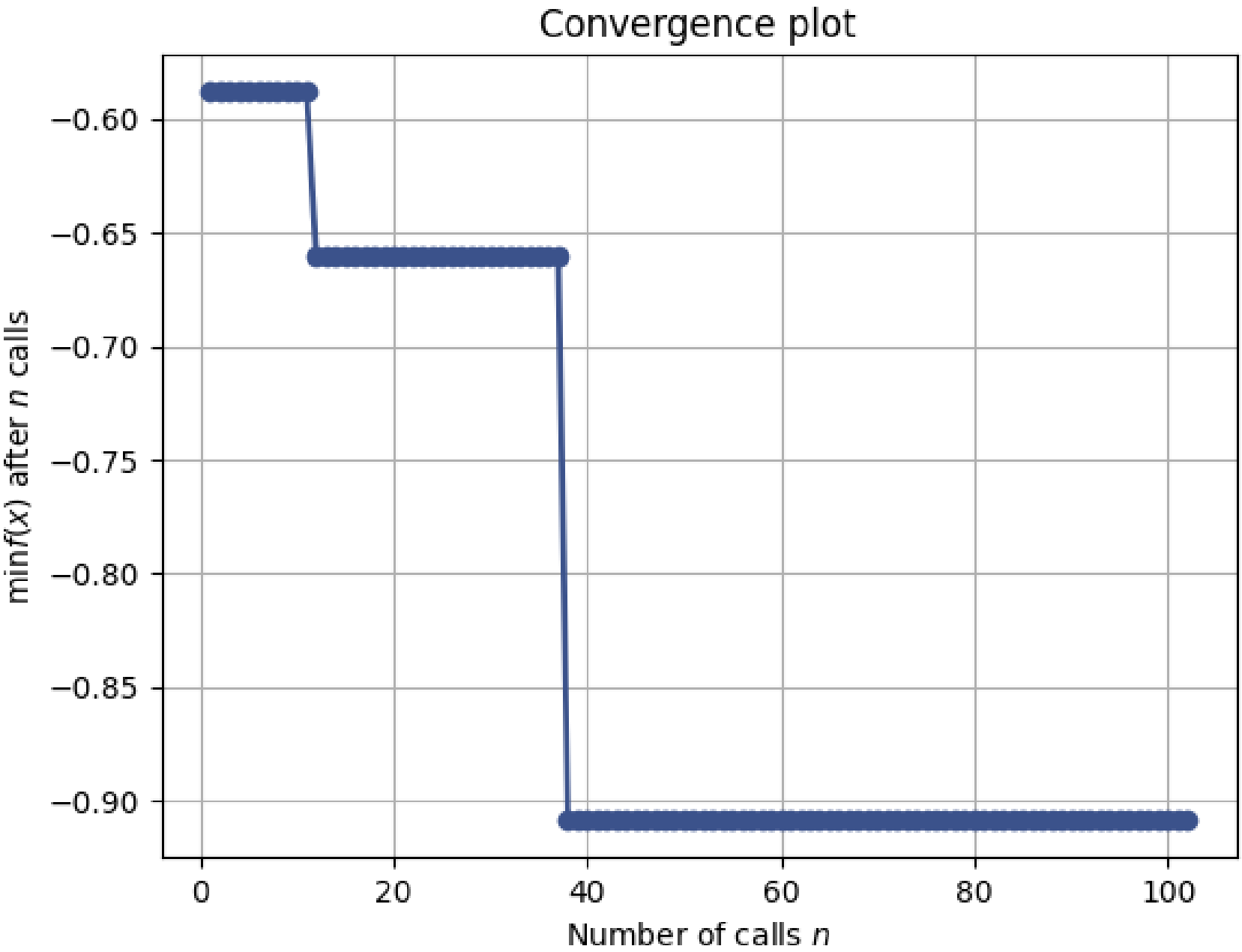
Ramki danych

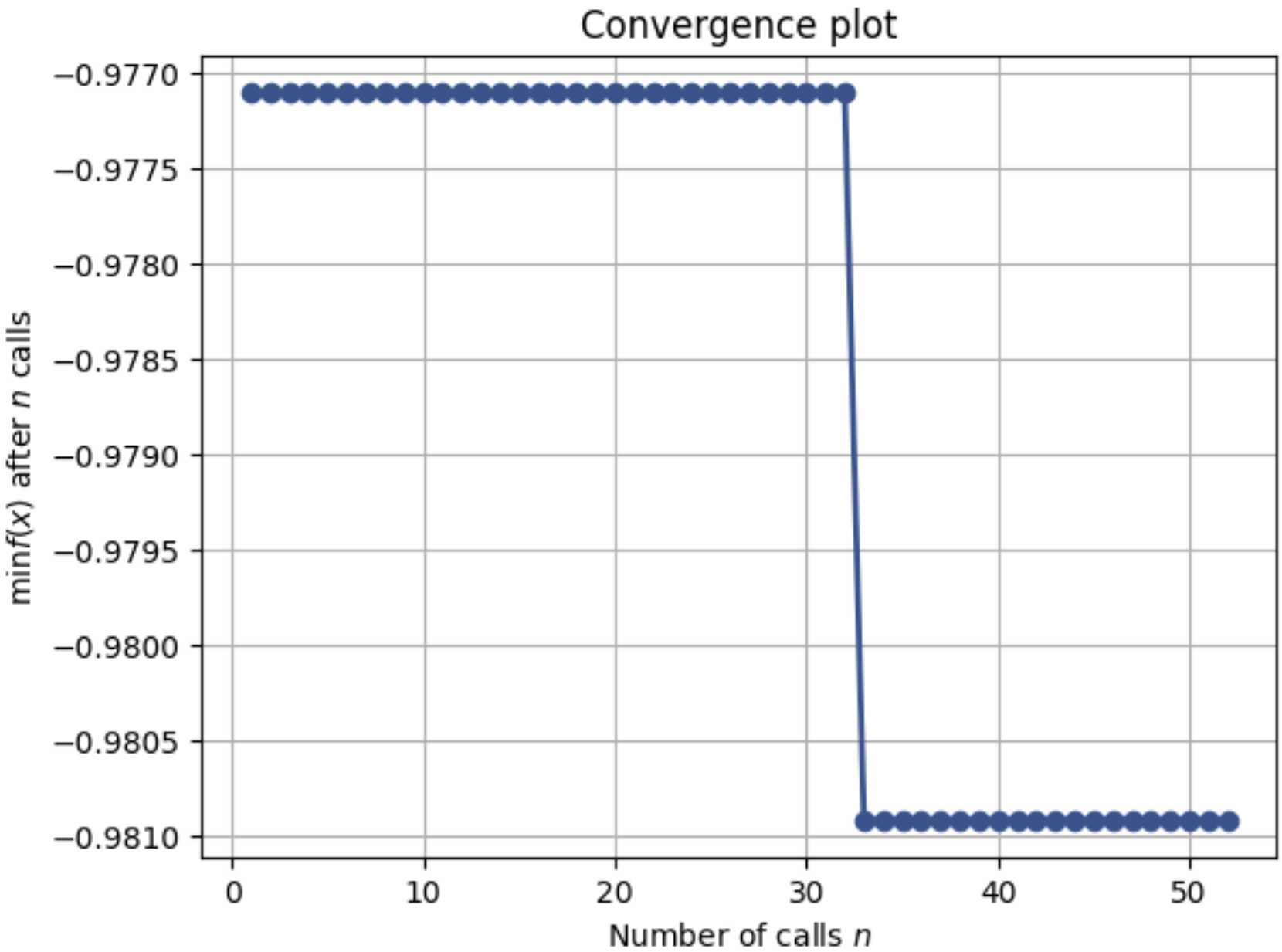
Parametry wybranych przez nas algorytmów optymalizowałyśmy na zbiorach danych:

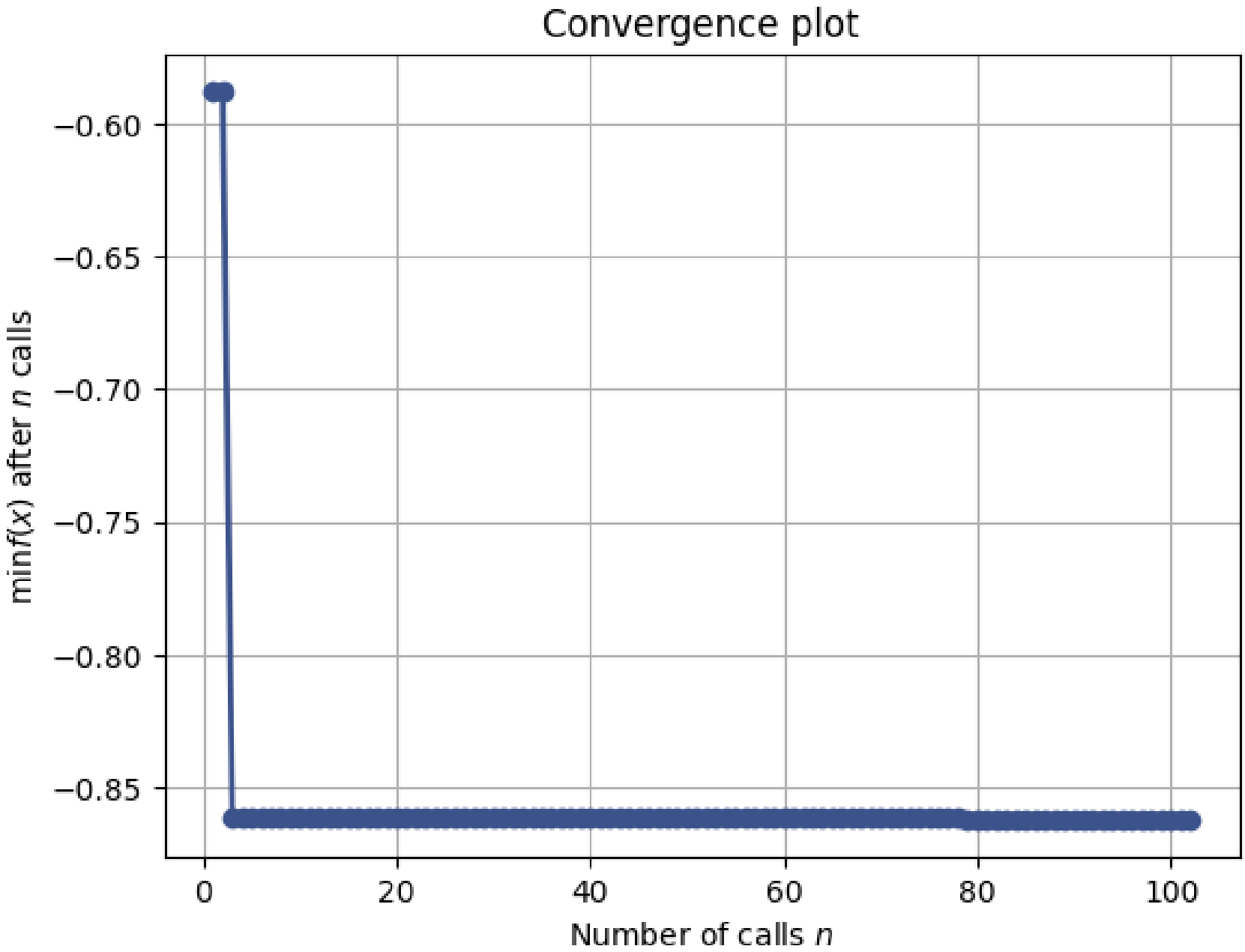
- Titanic,
- SpeedDating,
- Credit-g,
- Diabetes

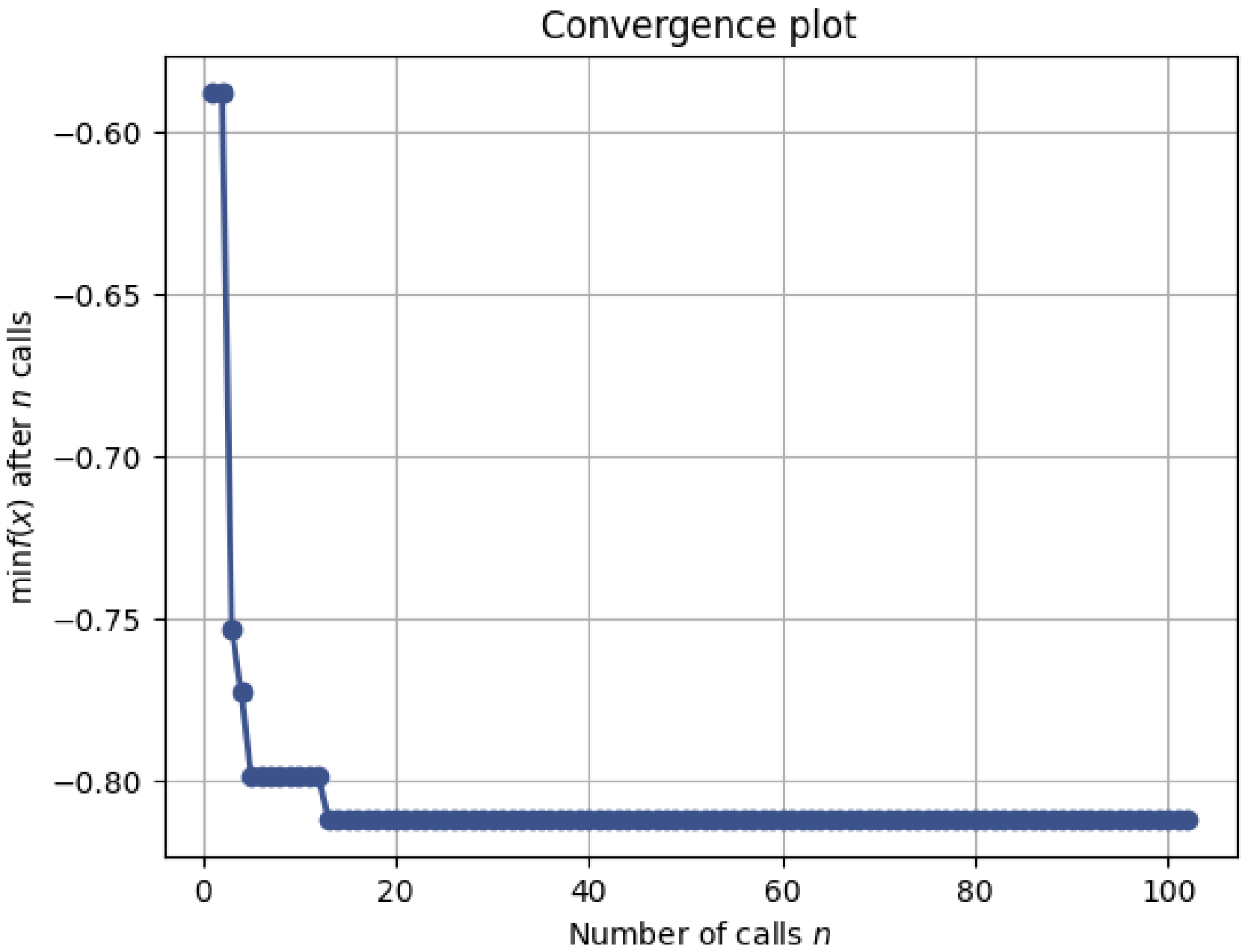
Hiperparametry oraz accuracy score				
Zbiór danych	**Metoda samplingu**	{*'n_estimators'*, *'min_samples_split'*, *'min_samples_leaf'*, *'max_features'*, *'max_depth'*}	**Score**	Zbieżność
Titanic	RandomizedSearchCV	{37, 8, 1, 12, 18}	0.7354568238779	-
	Bayes	{34, 2, 1, 13, 20}	0.9083969465649	38
SpeedDating	RandomizedSearchCV	{107, 2, 1, 11, 15}	0.8646669337874	-
	Bayes	{112, 18, 1, 13, 19}	0.8621718377083	3
Credit-g	RandomizedSearchCV	{45, 4, 3, 2, 8}	0.72	-
	Bayes	{117, 17, 1, 2, 20}	0.745	21
Diabetes	RandomizedSearchCV	{50, 6, 7, 4, 10}	0.7720778355325	-
	Bayes	{119, 17, 1, 2, 20}	0.8116883116883	16

Tabela 1: Optymalne kombinacje hiperparametrów dla zbiorów danych 1-4 korzystając z RandomizedSearchCV i Optymalizacji Bayesowskiej z algorytmem RandomForest.


Convergence plot
−0.60
−0.65
−0.70
−0.75
−0.80
−0.85
−0.90
min f(x) after n calls
0
20
40
60
80
100
Number of calls n


Convergence plot
−0.9770
−0.9775
−0.9780
−0.9785
−0.9790
−0.9795
−0.9800
−0.9805
−0.9810
min f(x) after n calls
0
10
20
30
40
50
Number of calls n

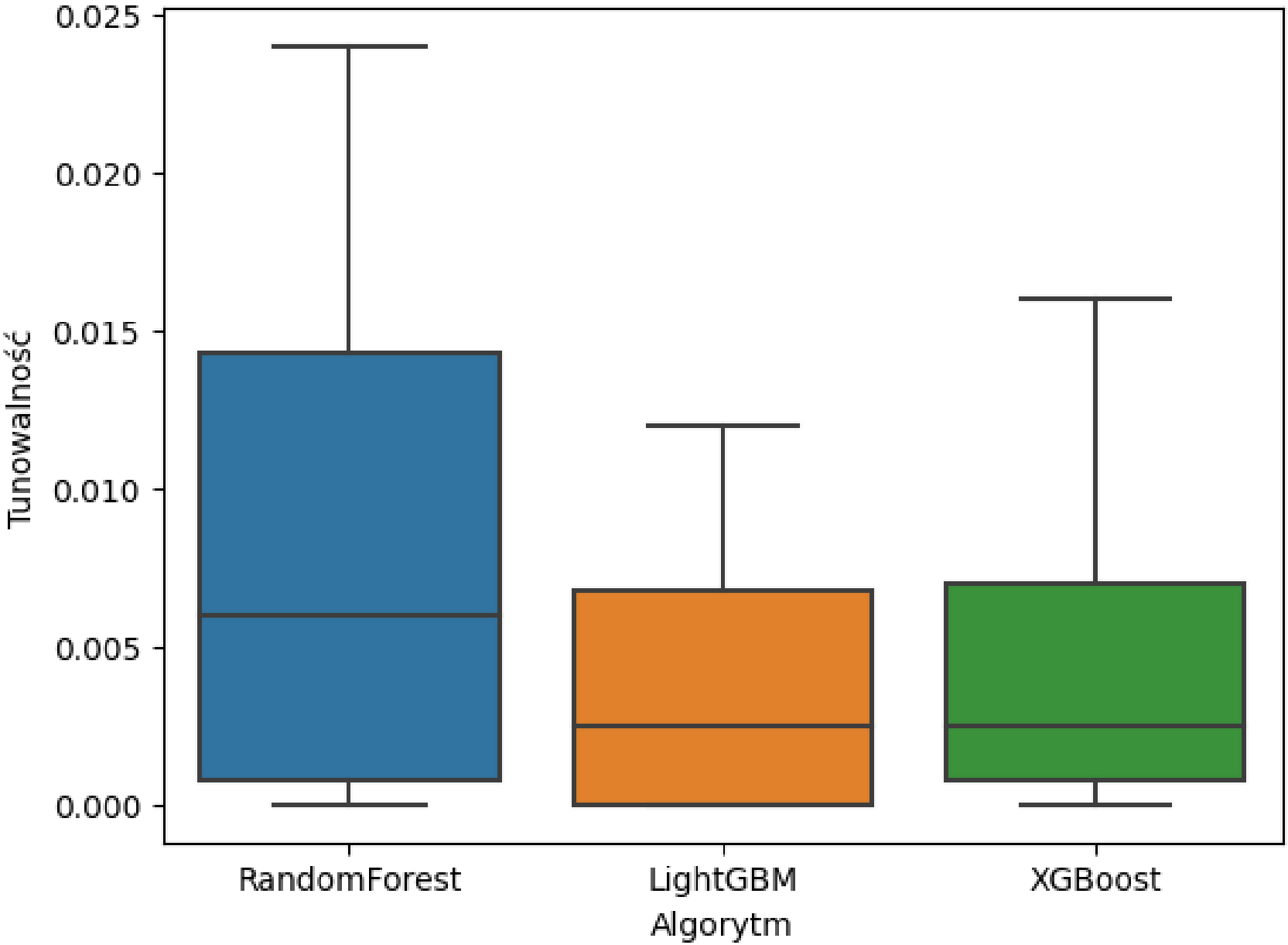

Convergence plot
−0.60
−0.65
−0.70
−0.75
−0.80
−0.85
min f(x) after n calls
0
20
40
60
80
100
Number of calls n


Convergence plot
−0.60
−0.65
−0.70
−0.75
−0.80
min f(x) after n calls
0
20
40
60
80
100
Number of calls n

Tunowalność algorytmów

Tunowalność		
Algorytm	Miara Tunowalności	Średnia tunowalność
RandomForest	$d^{(1)} = R^{(1)}(\theta^*) - R^{(1)}(\theta^{(1)*}) = -0.735 + 0.735 = 0$	0.00900
	$d^{(2)} = R^{(2)}(\theta^*) - R^{(2)}(\theta^{(2)*}) = -0.864 + 0.865 = 0.001$	
	$d^{(3)} = R^{(3)}(\theta^*) - R^{(3)}(\theta^{(3)*}) = -0.696 + 0.72 = 0.024$	
	$d^{(4)} = R^{(4)}(\theta^*) - R^{(4)}(\theta^{(4)*}) = -0.761 + 0.772 = 0.011$	
LightGBM	$d^{(1)} = R^{(1)}(\theta^*) - R^{(1)}(\theta^{(1)*}) = -0.975 + 0.975 = 0$	0.00425
	$d^{(2)} = R^{(2)}(\theta^*) - R^{(2)}(\theta^{(2)*}) = -0.863 + 0.875 = 0.012$	
	$d^{(3)} = R^{(3)}(\theta^*) - R^{(3)}(\theta^{(3)*}) = -0.708 + 0.708 = 0$	
	$d^{(4)} = R^{(4)}(\theta^*) - R^{(4)}(\theta^{(4)*}) = -0.743 + 0.748 = 0.005$	
XGBoost	$d^{(1)} = R^{(1)}(\theta^*) - R^{(1)}(\theta^{(1)*}) = -0.975 + 0.976 = 0.001$	0.00525
	$d^{(2)} = R^{(2)}(\theta^*) - R^{(2)}(\theta^{(2)*}) = -0.865 + 0.869 = 0.004$	
	$d^{(3)} = R^{(3)}(\theta^*) - R^{(3)}(\theta^{(3)*}) = -0.698 + 0.698 = 0$	
	$d^{(4)} = R^{(4)}(\theta^*) - R^{(4)}(\theta^{(4)*}) = -0.743 + 0.759 = 0.016$	

Tabela 4: Tunowalność dla algorytmów.


0.025
0.020
0.015
0.010
0.005
0.000
Tunowalność
RandomForest
LightGBM
XGBoost
Algorytm

Dziękujemy za uwagę!