Analiza tunowalności algorytmów

Dominik Kędzierski
Grzegorz Zakrzewski

Zbiory danych

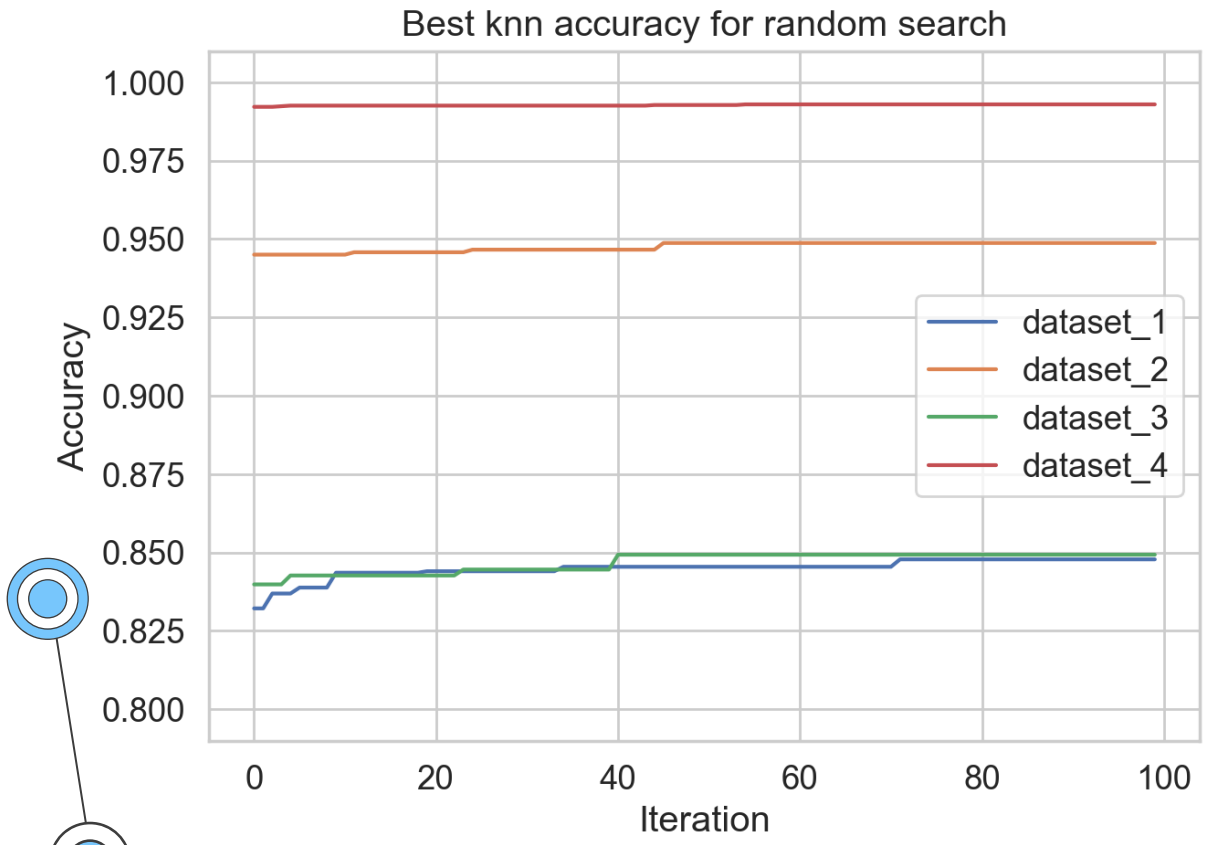
Zbiór danych	*ID*	*Liczba obserwacji*	*Liczba zmiennych*
KC1 software defect prediction	1067	2 109	22
Amazon employee access	4135	32 769	9
QSAR biodegradation	1494	1 055	41
Satellite	40900	5100	36

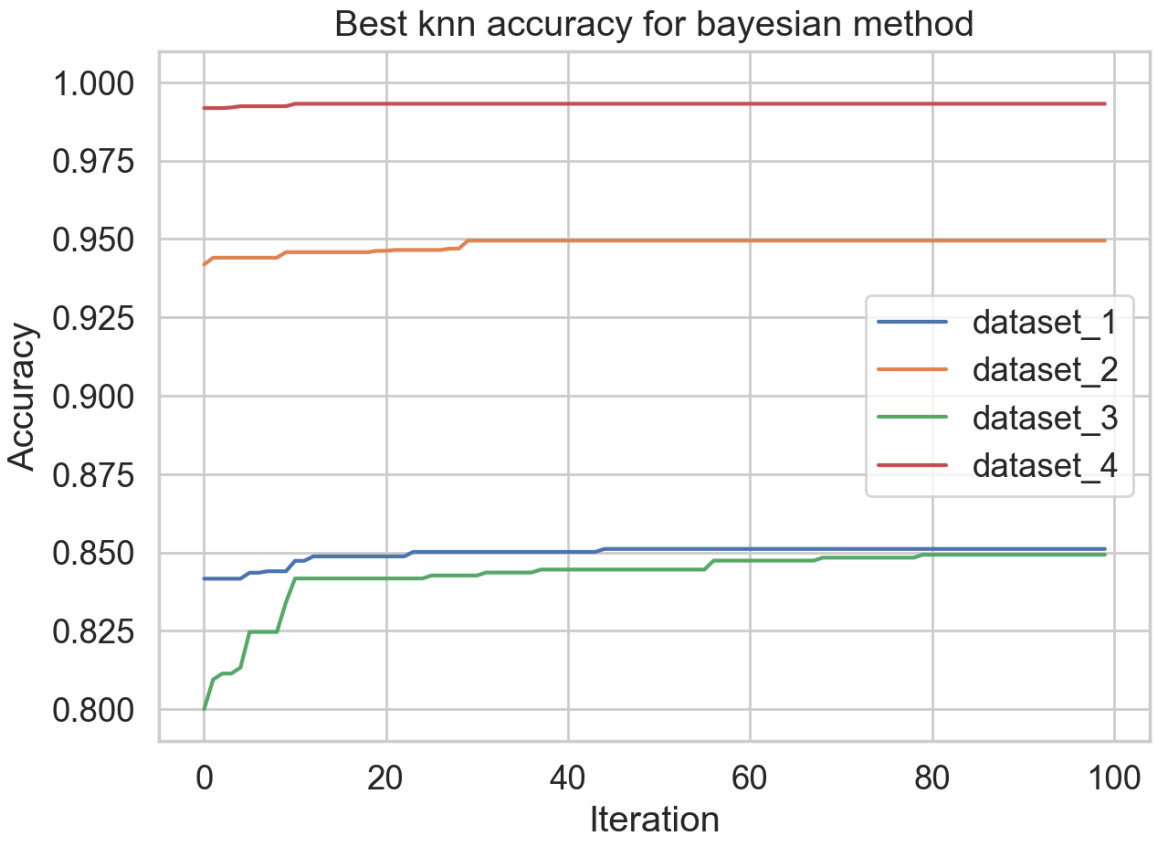
- Binarna zmienna celu
- Nieduże zbiory danych
- Nie zawierają braków danych

Algorytmy i hiperparametry

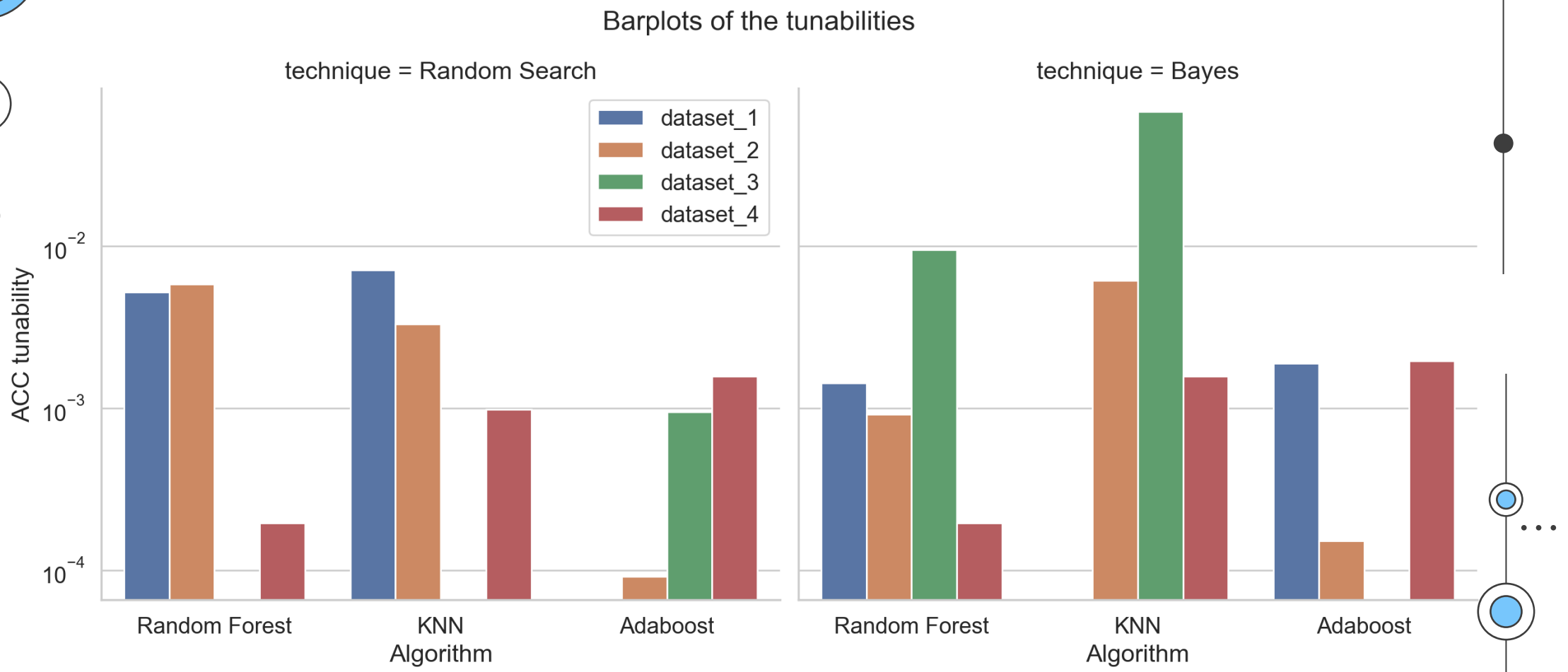
Algorytm	*Hiperparametr*	*Zakres wartości*
Random Forest	n_estimators	{ 50, …, 300 }
	critertion	{ gini, entropy, log_loss }
	max_depth	{ 1, …, 50 }
	min_samples_split	{ 2, …, 20 }
	max_features	{ sqrt, log2, None }
K-Nearest Neighbors	n_neighbors	{ 1, …, 30 }
	p	[-1.0, 2.0]
	weights	{ uniform, distance }
AdaBoost	n_estimators	{ 50, …, 500 }
	learning_rate	[0.01, 2.0] (log-scale)
	algorithm	{ SAMME.R, SAMME }

Analiza zbieżności algorytmów





Analiza tunowalności algorytmów



Dziękujemy za uwagę