

Analiza tunowalności wybranych algorytmów uczenia maszynowego

Michał Syrkiewicz, Jan Taran, Piotr Wysocki

17 listopada 2025

1 Wybrane algorytmy i zbiory danych

Użyte w projekcie algorytmy: Random Forest Classifier, LightGBM, SVM z jądrem RBF
Użyte zbiory danych: Bank Marketing, Titanic, Heart Failure Prediction , Gym Members Exercise
Wybrana metryka wydajności: Brier Score (minimalizowana jako ryzyko)

2 Metodologia

2.1 Zakresy hiperparametrów

Zakresy hiperparametrów zostały określone na podstawie literatury (Probst et al., 2019), (Probst et al., 2018) oraz dokumentacji bibliotek. Poniższa tabela przedstawia wybrane zakresy:

Tabela 1: Zakresy hiperparametrów dla badanych algorytmów

Algorytm	Hiperparametr	Zakres	Transformacja
Random Forest	n_estimators	[100, 1500)	-
	criterion	['log_loss', 'entropy', 'gini']	-
	max_depth	[10, 15, 20, None]	-
	min_samples_split	[2, 20)	-
	min_samples_leaf	[1, 10)	-
	max_samples	[0.5, 0.7, 0.9, None]	-
	max_features	['sqrt', 0.1, 0.25, 0.33]	-
LightGBM	num_leaves	[15, 128]	-
	max_depth	[3, 15]	-
	learning_rate	[0.01, 0.3]	log
	n_estimators	[100, 1500]	-
	min_data_in_leaf	[10, 200]	-
	lambda_l1	[1e-4, 10]	log
	lambda_l2	[1e-4, 10]	log
SVM	C	[0.001, 1000]	log
	gamma	[0.0001, 10]	log

2.2 Metody tuningu

Dla obu metod przeprowadziliśmy po 100 iteracji . W tuningu zastosowana została 5-krotna CV.

1. **Random Search:** Jednostajne losowanie z określonej siatki hiperparametrów
2. **Bayesian Optimization:** Z wykorzystaniem pakietu Optuna

2.3 Miary tunowalności

Zgodnie z definicjami z artykułu (Probst et al., 2019), tunowalność została wyznaczona dla metryki **Brier score**. **Optymalna domyślna konfiguracja θ^* :**

$$\theta^* := \arg \min_{\theta \in \Theta} g(R^{(1)}(\theta), \dots, R^{(m)}(\theta)) \quad (1)$$

gdzie g to funkcja agregująca. W naszym przypadku zastosowano **medianę** po wartościach każdego hiperparametru dla wszystkich zbiorów dla danego algorytmu, ponieważ niektóre zbiory danych wykazywały

znaczne odstępstwa, a liczba zbiorów była parzysta — umożliwia to uśrednienie dwóch śródkowych wartości i zapewnia bardziej stabilny wynik niż średnia. **Tunowalność algorytmu** na zbiorze j :

$$d^{(j)} := R^{(j)}(\theta^*) - R^{(j)}(\theta^{(j)*}) \quad (2)$$

gdzie $R^{(j)}$ oznacza wynik Brier score na zbiorze j , θ^* — optymalną konfigurację domyślną, a $\theta^{(j)*}$ — optymalną konfigurację specyfczną dla danego zbioru.

3 Wyniki eksperymentów

3.1 Nowe domyślne konfiguracje

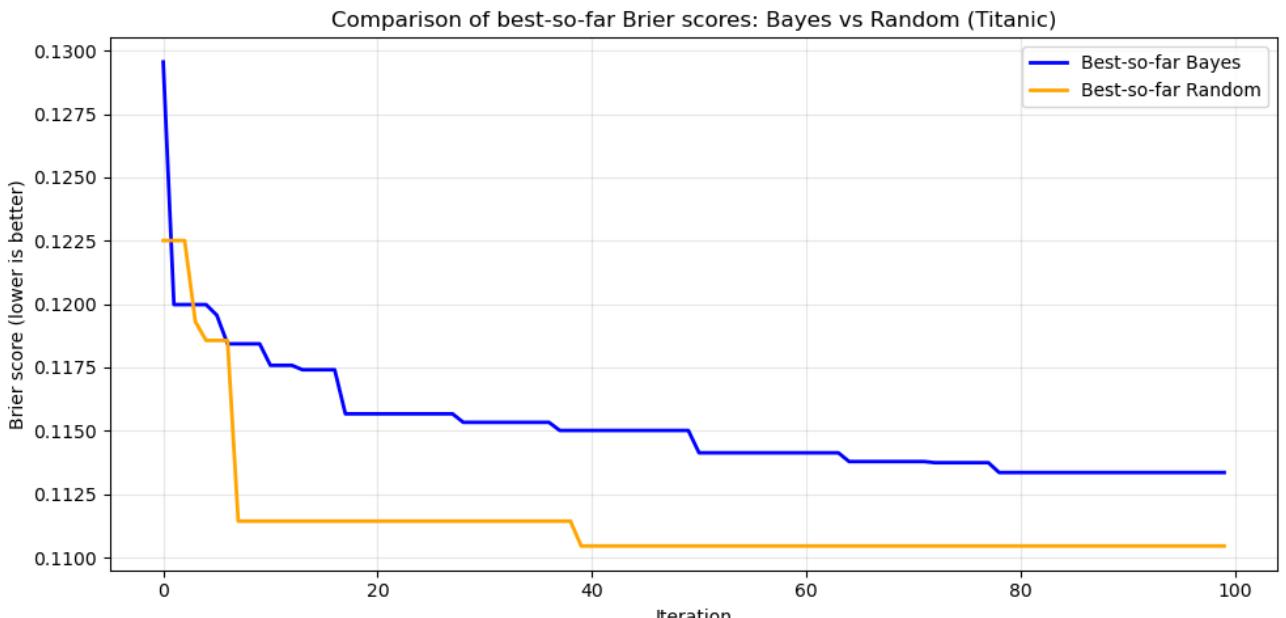
Nowa domyślna konfiguracja dla każdego algorytmu została wyznaczona jako mediana z wartości numerycznych lub moda z wartości kategorycznych z tych samych parametrów, które uzyskały najlepsze wyniki Brier Score.

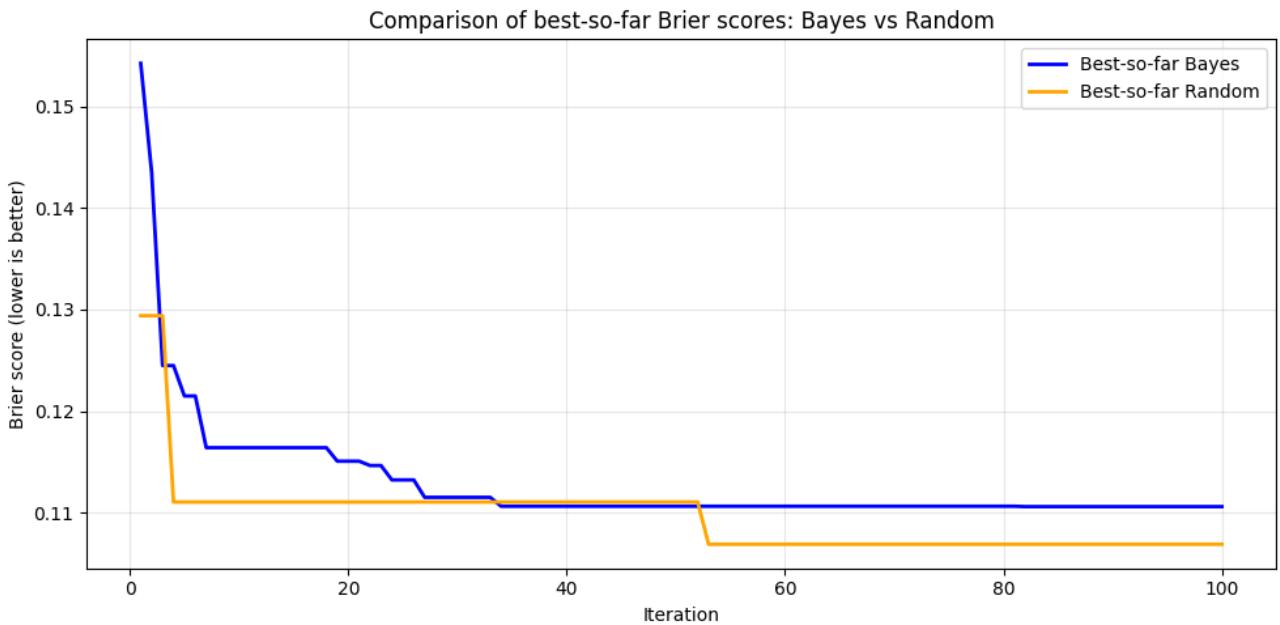
Tabela 2: Porównanie domyślnych i nowych optymalnych konfiguracji

Algorytm	Hiperparametr	Default (sklearn)	Nowy default
Random Forest	n_estimators	100	1347
	criterion	gini	entropy
	max_depth	None	None
	min_samples_split	2	5
	min_samples_leaf	1	1
	max_samples	None	None
	max_features	sqrt	0.25
LightGBM	num_leaves	31	83
	learning_rate	0.1	0.0251
	n_estimators	100	270
	max_depth	-1	4
	min_data_in_leaf	20	54
	lambda_l1	0.0	0.0166
	lambda_l2	0.0	0.7994
SVM (rbf)	C	1.0	6.290
	gamma	scale ^a	0.0136

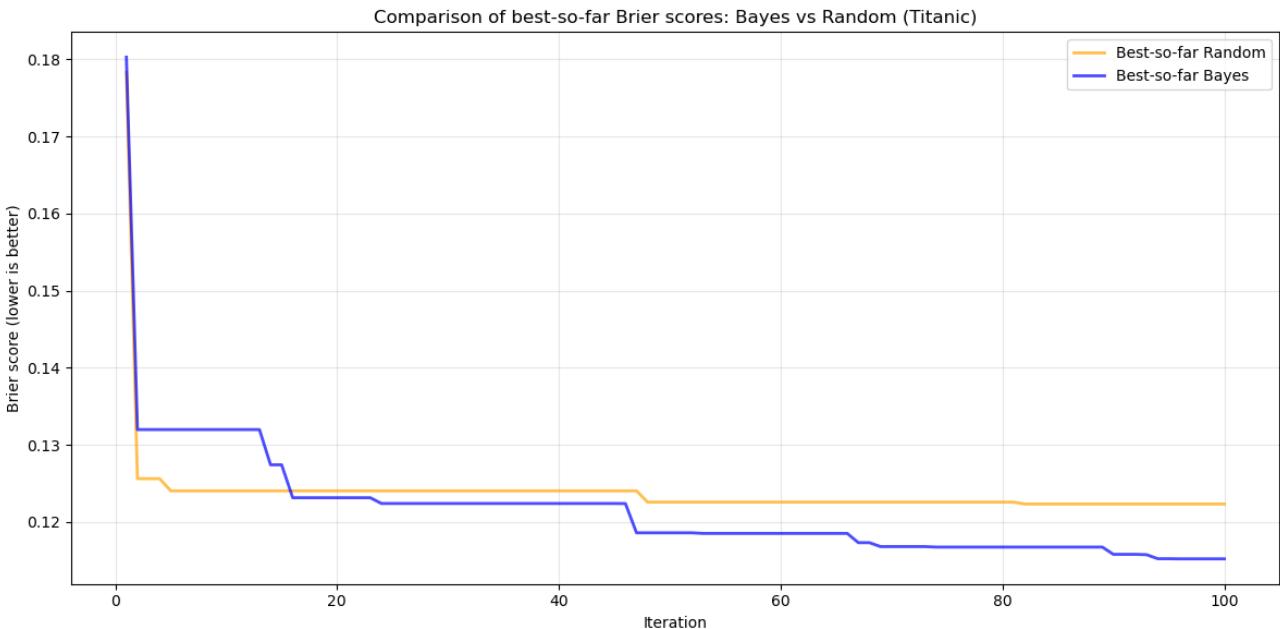
^a Dla opcji scale: $\gamma = \frac{1}{n_{\text{features}} \cdot \text{Var}(X)}$

3.2 Analiza zbieżności dla obu metod równocześnie





Rysunek 1: LightGBM – zbieżność



Rysunek 2: SVM – zbieżność

3.3 Tunability algorytmów

Tabela 3: Porównanie poprawy metryki Brier score po wybraniu nowego 'defaultowego' zbioru hiperparametrów względem domyślnych wartości dla każdego z algorytmów.

Dataset	Random Forest		LightGBM		SVM	
	Improvement	Relative (%)	Improvement	Relative (%)	Improvement	Relative (%)
Bank data	+0.0023	+1.2%	-0.0019	-1.0%	+0.001396	+0.737%
Gym data	+0.0035	+1.9%	+0.0226	+11.2%	+0.001701	+1.009%
Heart disease	+0.0023	+2.2%	+0.0052	+5.3%	-0.0029	-2.926%
Titanic	+0.0007	+0.5%	+0.0218	+12.2%	-0.0027	-1.599%

Ranking tunowalności algorytmów (średnia wartość d):

1. LightGBM: $d = 0.0119 \pm 0.0122$
2. RandomForest: $d = 0.0022 \pm 0.00098$
3. SVM: $d = -0.000641 \pm 0.00219$

3.4 Porównanie technik losowania - bias sampling

Test statystyczny:

- Test: Wilcoxon dla par sparowanych. Został przeprowadzony na 12 parach Brier Score (dla wszystkich kombinacji 3 algorytmów i 4 zbiorów danych). Hipoteza zerowa zakłada równość median.
- p-value: 0.5693
- Wnioski: Obliczona p-wartość jest bardzo wysoka, więc na podstawie naszych danych nie możemy stwierdzić, że istnieje istotna różnica między jakością obu badanych technik losowania hiperparametrów.

3.5 Wpływ wielkości zbioru danych na wyniki tunowania

Analiza została przeprowadzona na zbiorze *Bank data*, ponieważ był on najbardziej liczny. Dla każdego algorytmu obliczono różnicę w wartości metryki Brier score (Δ Brier) pomiędzy modelami po tuningu a modelem tunowanym na pełnym zbiorze danych.

Tabela 4: Wpływ wielkości zbioru treningowego na różnicę metryki Brier score (Δ Brier).

Algorytm	25%	50%	75%
LightGBM	0.0184	0.0093	0.0068
RandomForest	0.0106	0.0106	0.0078
SVM	0.0080	0.0058	0.0020

Wnioski:

- Różnice maleją w sposób zbliżony do liniowego - im mniejszy zbiór, tym większa różnica Δ Brier, co wskazuje na gorszą efektywność tuningu dla mniejszych zbiorów. Sugeruje to, że tunowalność jest bardziej wyraźna w dużych zbiorach, gdzie hiperparametry mogą lepiej wychwytywać złożone wzorce.

4 Wnioski

1. **Tunability algorytmów:** Analiza wyników Brier score pokazuje, że największy potencjał poprawy poprzez tuning mają algorytmy LightGBM (średnia poprawa $d = 0.0119 \pm 0.0122$), natomiast Random Forest (średnia $d = 0.0022 \pm 0.00098$) i SVM (średnia $d = -0.000641 \pm 0.00219$) wykazują minimalne lub mieszane efekty.
2. **Nowe domyślne konfiguracje:** Wprowadzenie nowych wartości domyślnych ma ograniczony wpływ na poprawę wyników w przypadku Random Forest i SVM, natomiast dla LightGBM zmiany parametrów pozwalają na umiarkowaną poprawę wydajności, szczególnie na zbiorach Titanic i Gym.
3. **Bias sampling:** Na podstawie naszych danych nie ma statystycznie przekonujących dowodów na jego występowanie.

Literatura

Probst, P., Boulesteix, A.-L., and Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. *Journal of Machine Learning Research*.

Probst, P., Wright, M. N., and Boulesteix, A.-L. (2018). Hyperparameters and tuning strategies for random forest. *WIREs Data Mining Knowl Discov* 2019.