



Tunowalność algorytmów

Agata Kopyt, Zuzanna Kotlińska

Tunowalność

Elastyczność w dostosowywaniu parametrów modelu w celu optymalizacji jego wydajności. Im większa tunowalność algorytmu, tym łatwiej dostosować go do różnych zestawów danych i problemów poprzez modyfikację hiperparametrów.

A stylized, light blue leaf graphic with visible veins, positioned in the top right corner of the slide.

Eksperyment

Przeanalizowanie tunowalności
hiperparametrów dla trzech różnych
algorytmów uczenia maszynowego
oraz czterech różnych zbiorów danych.

Dane

- **Źródło: OpenML,**
- **Użyte w eksperymencie zbiory:**
 1. kc2,
 2. fri-c2-500-25,
 3. breast-2,
 4. diabetes.
- **Wszystkie wykorzystane zbiory danych służą do zadania klasyfikacji.**





Modele

- **Random Forest,**
- **Elastic Net,**
- **Decision Tree.**

Metody samplingu

Random search

metoda oparta na wyborze punktów z rozkładu jednostajnego, w której wartości hiperparametrów są wybierane losowo z siatki.

Bayes Optimization

metoda oparta na technice bayesowskiej, wykorzystująca model probabilistyczny do przewidywania, które kombinacje hiperparametrów są najbardziej optymalne.



Siatki hiperparametrów

Hiperparametr	Typ	Wartości
n_estimators	integer	[1, 2000]
criterion	string	gini, entropy
max_depth	integer	[3, 10]
min_samples_split	integer	[2, 10]
min_samples_leaf	integer	[1, 10]
bootstrap	logical	True, False
max_samples	numeric	[0, 1]

Tabela: Zestaw hiperparametrów dla Random Forest

Siatki hiperparametrów

Hiperparametr	Typ	Wartości
alpha	numeric	2^x dla $x \in [-10, 10]$
l1_ratio	numeric	[0, 1]

Tabela: Zestaw hiperparametrów dla Elastic Net

Siatki hiperparametrów

Hiperparametr	Typ	Wartości
max_depth	integer	[1, 30]
max_features	string	None, log2, sqrt
criterion	string	gini, entropy
splitter	string	best, random
min_samples_split	integer	[1, 60]
min_samples_leaf	integer	[1, 60]

Tabela: Zestaw hiperparametrów dla Decision Tree



Trening i ewaluacja

- Znaleziono zostało 100 różnych zestawów hiperparametrów dla każdego zbioru danych, metody samplingu i modelu uczenia maszynowego.
- Metryka do ewaluacji: ROC AUC.
- Wzór, według którego liczona jest tunowalność dla danego algorytmu, dla j-tego zbioru danych:

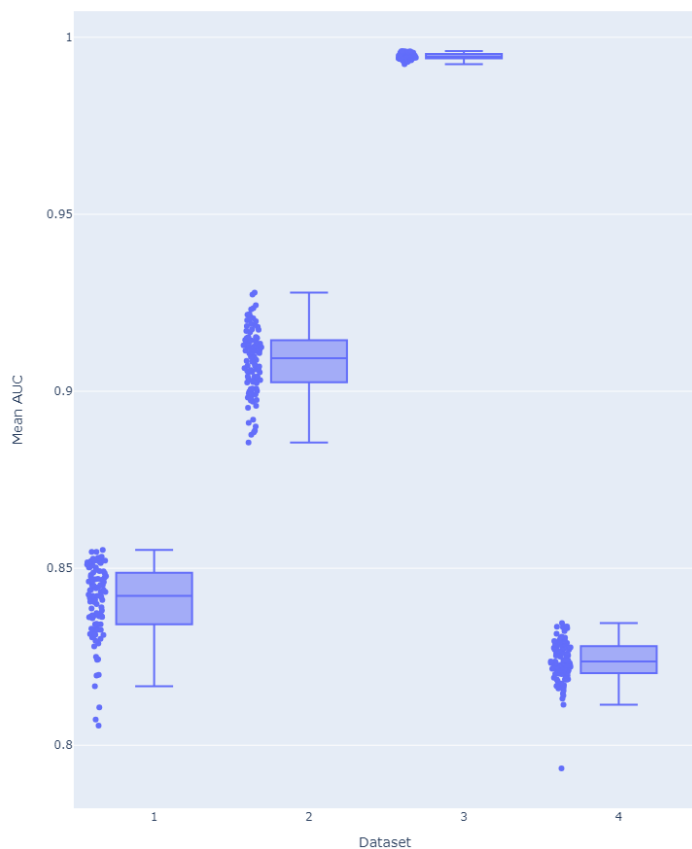
$$d^{(j)} := R^{(j)}(\theta^*) - R^{(j)}(\theta^{(j)*})$$

Wyniki

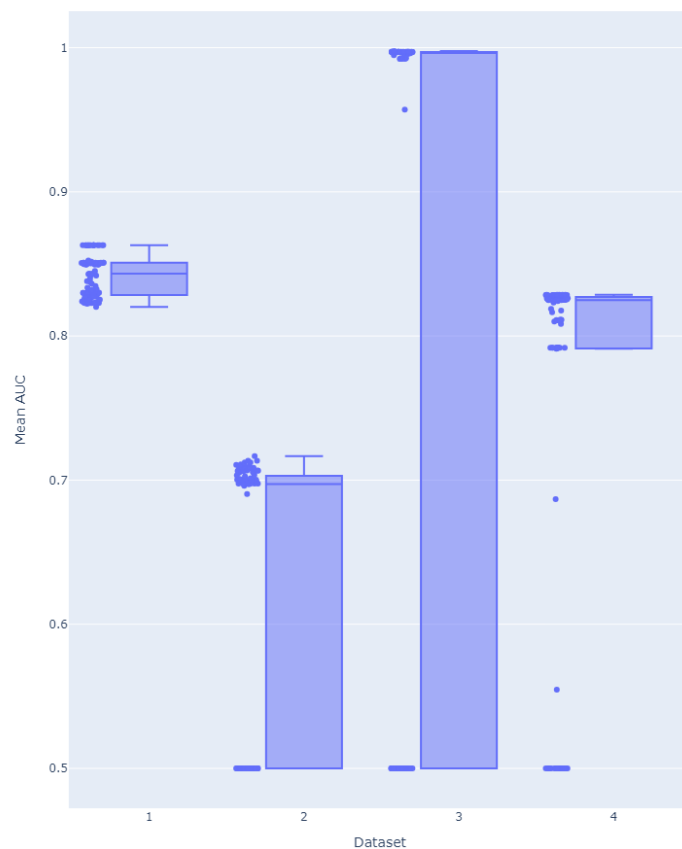


Random Search - AUC

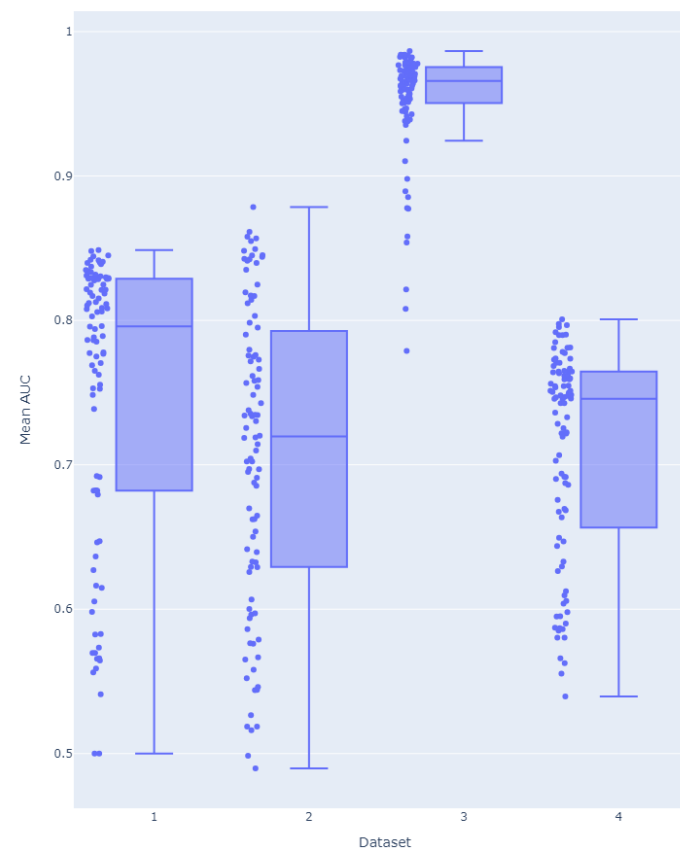
RS: Average performance for Random Forest



RS: Average performance for Elastic Net



RS: Average performance for Decision Tree

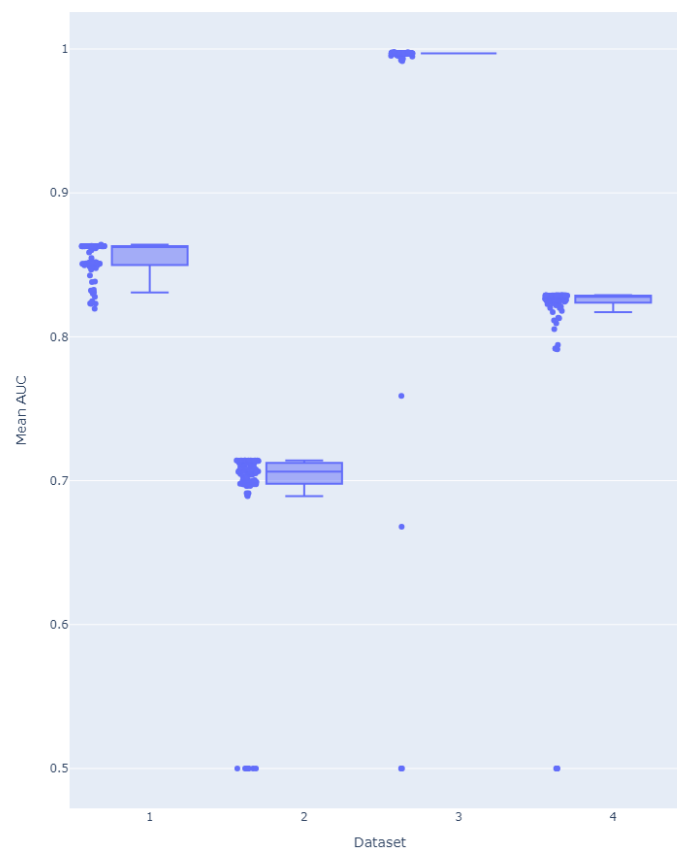


Bayesian Optimization - AUC

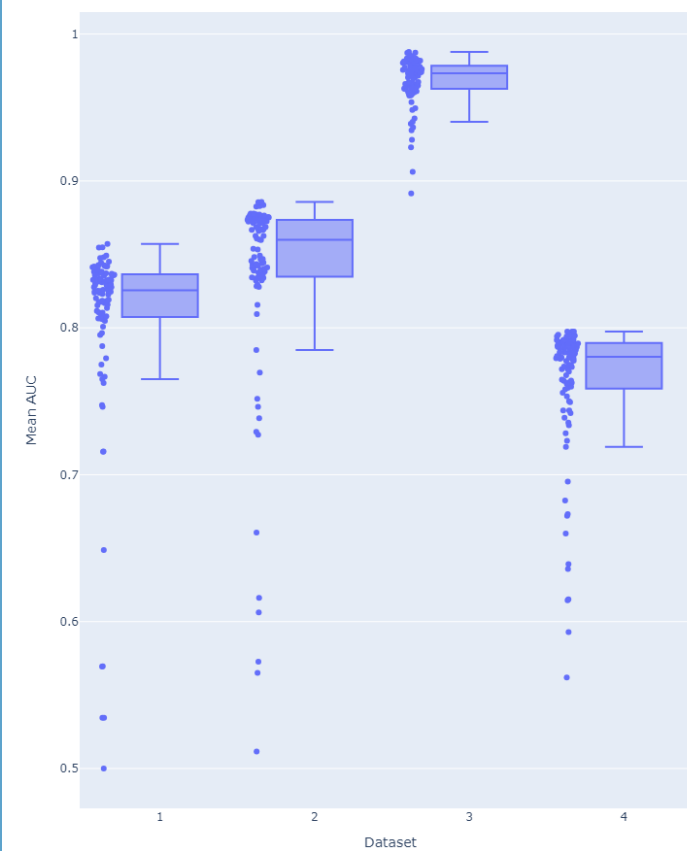
BO: Average performance for Random Forest



BO: Average performance for Elastic Net

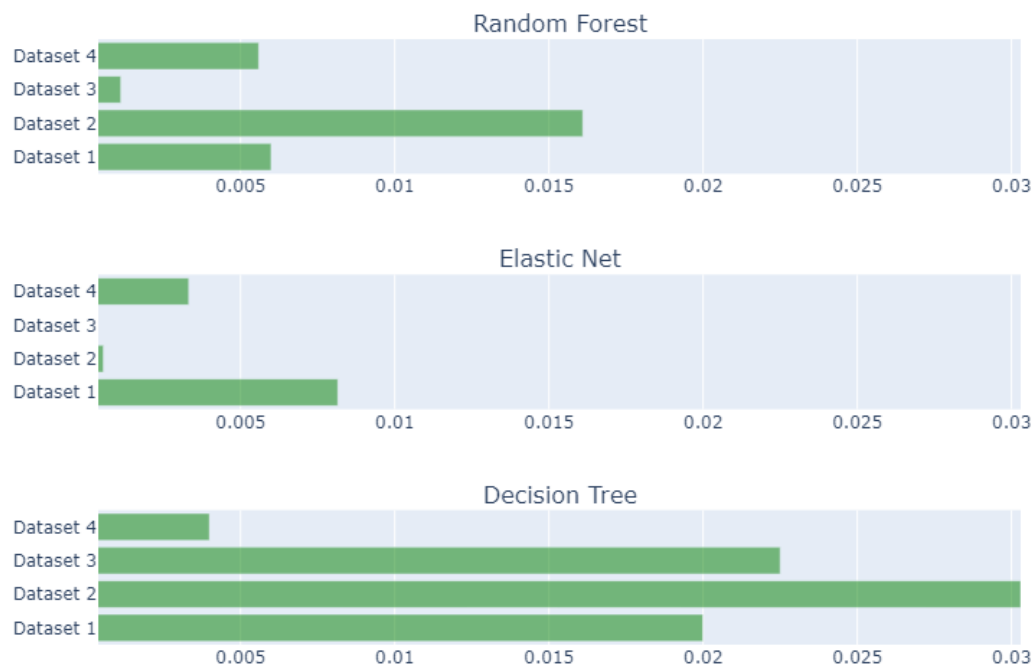


BO: Average performance for Decision Tree



Tunowalność

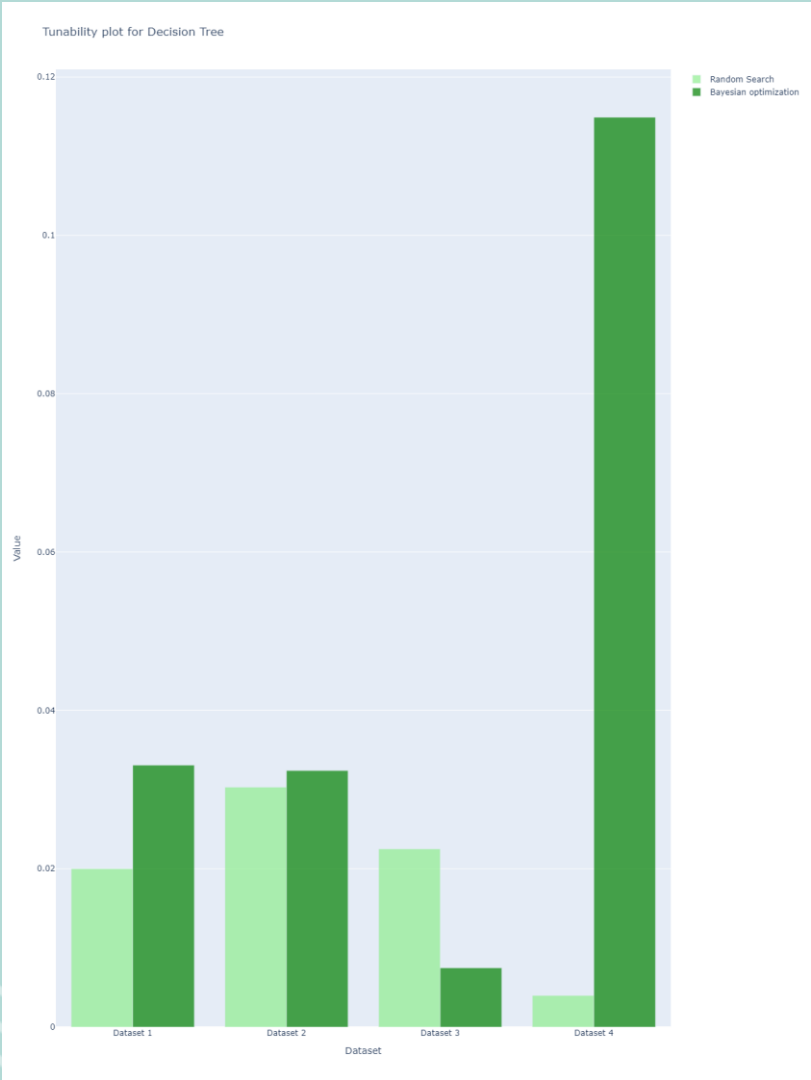
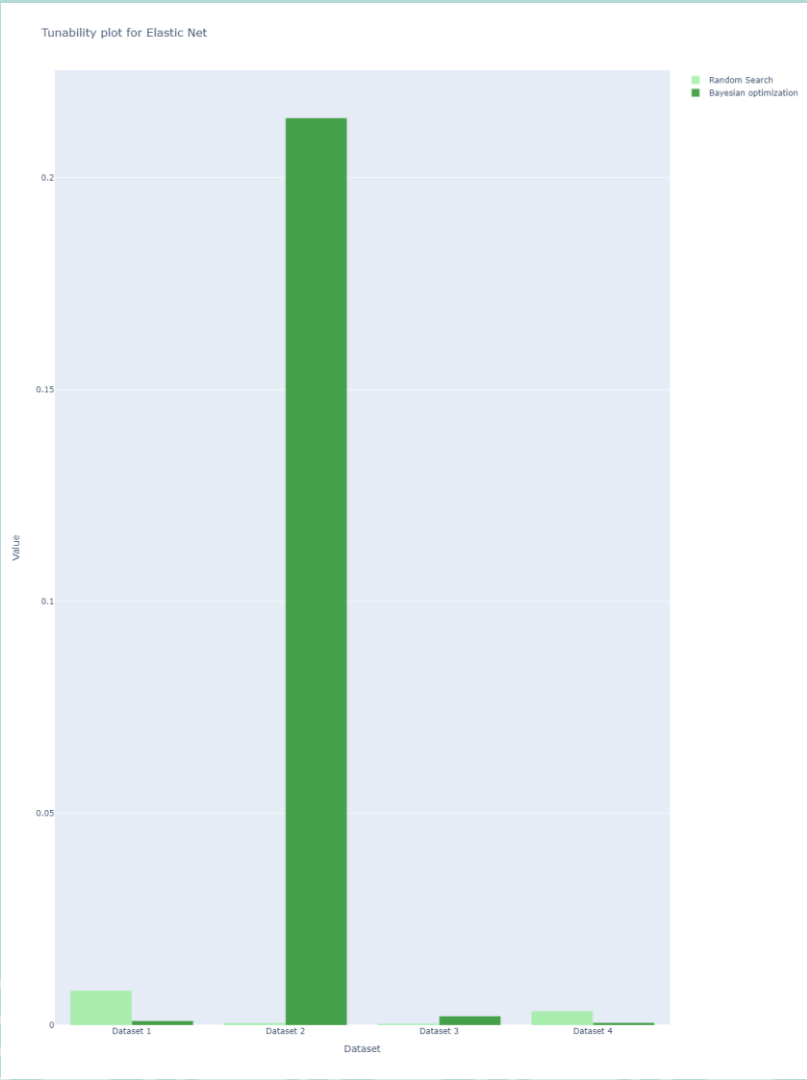
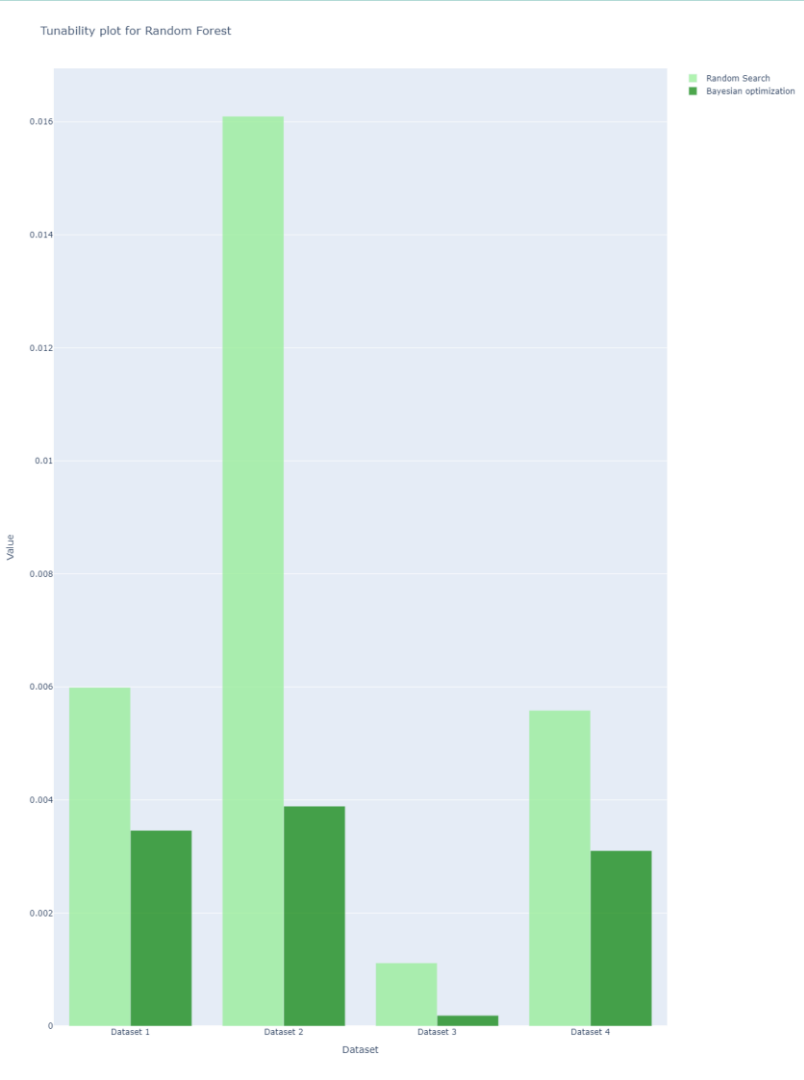
Tunability using Random search



Tunability using Bayes optimisation



Tunowalność



Wnioski

- Decision Tree okazał się najbardziej tunowalnym algorytmem,
- Korzystając z defaultu dla Elastic Net, występują szczególne przypadki, dla których default znacząco pogorszy wyniki,
- Korzystanie z defaultu dla Random Forest i Elastic Net ma sens, jeśli niewielka poprawa dokładności modelu nie jest istotna,
- Nie można jednoznacznie stwierdzić wpływu metody strojenia parametrów na tunowalność algorytmu.



Źródła

Testowane zbiory danych pochodzą z OpenML:

- <https://www.openml.org>

Siatki hiperparametrów oraz wzór na tunowalność wykorzystany w eksperymencie pochodzi z poniższego artykułu:

- <https://jmlr.org/papers/volume20/18-444/18-444.pdf>



Dziękujemy za uwagę