

---

# Raport z Realizacji Projektu Badawczego Automatyczne Uczenie Maszynowe

## Optymalizacja i Porównanie Metod Strojenia Hiperparametrów dla Wybranych Algorytmów z Wykorzystaniem Technik Samplingu Rok akademicki 2024/2025

Wersja 1.0  
Andrii Voznesenskyi

Opiekun naukowy:  
mgr Anna Kozak

19 listopada 2024

# Spis treści

<b>1</b>	<b>Wstęp</b>	<b>4</b>
<b>2</b>	<b>Zbiory danych</b>	<b>4</b>
2.1	Przetwarzanie danych . . . . .	4
<b>3</b>	<b>Metodyka</b>	<b>4</b>
3.1	Wybrane Algorytmy . . . . .	4
3.2	Techniki Tunowania . . . . .	5
<b>4</b>	<b>Wyniki</b>	<b>5</b>
4.1	Tunowalność Algorytmów . . . . .	5
4.2	Zbieżność wyników i stabilizacja modeli . . . . .	6
4.3	Diagram Różnic Krytycznych . . . . .	7
<b>5</b>	<b>Wnioski</b>	<b>7</b>

## **Streszczenie**

Niniejszy raport prezentuje wyniki analizy tunowalności hiperparametrów dla czterech algorytmów uczenia maszynowego: XGBoost, Random Forest, ElasticNet i Gradient Boosting. Wykorzystano trzy techniki tunowania: Grid Search, Random Search oraz Optymalizację Bayesowską. Przeanalizowano 10 zbiorów danych pod kątem efektywności, stabilności oraz tunowalności algorytmów. Wyniki wskazują, że Optymalizacja Bayesowska była najbardziej efektywna, osiągając stabilne rezultaty przy mniejszej liczbie iteracji. XGBoost i Gradient Boosting wykazały największą tunowalność, co sugeruje ich podatność na poprawę wydajności dzięki odpowiedniej konfiguracji hiperparametrów.

# 1 Wstęp

Celem raportu jest analiza tunowalności hiperparametrów dla algorytmów XGBoost, Random Forest, ElasticNet oraz Gradient Boosting z wykorzystaniem metod *Grid Search*, *Random Search* i *Optymalizacji Bayesowskiej*. Tunowalność oznacza wrażliwość wyników modelu na zmiany konfiguracji hiperparametrów, co pozwala ocenić, jak dobrze dany algorytm może być dostosowany do danych.

Badanie obejmuje siedem dobrze przygotowanych zbiorów danych: *Breast Cancer*, *Iris*, *California Housing*, *Digits*, *Wine*, *Diabetes* oraz *Blood Transfusion*. Zbiory *Auto MPG* i *Auto Insurance* zostały odrzucone z analizy ze względu na ich niski stopień przygotowania, brak konwersji zmiennych ciągłych na kategorię oraz liczne brakujące wartości.

Metryki dostępne w bibliotekach takich jak `scikit-learn` zostały zachowane i wykorzystane w ocenie efektywności modeli. Przeprowadzono także test F, aby statystycznie porównać metody tunowania pod względem wyników na różnych zbiorach danych i modelach.

## 2 Zbiory danych

Do analizy wykorzystano 10 zróżnicowanych zbiorów danych, obejmujących zadania klasyfikacyjne i regresyjne:

**Breast Cancer** (*scikit-learn*) to dane klasyfikacyjne dotyczące diagnostyki raka piersi, zawierające 30 cech histopatologicznych. **Iris** (*scikit-learn*) to klasyczny zbiór klasyfikacyjny opisujący cechy trzech gatunków irysów. **California Housing** (*scikit-learn*) obejmuje dane regresyjne dotyczące cen domów w Kalifornii z uwzględnieniem cech demograficznych. **Wine** (*scikit-learn*) to dane klasyfikacyjne dotyczące jakości wina na podstawie analizy chemicznej. **Digits** (*scikit-learn*) zawiera dane klasyfikacyjne z obrazami cyfr w formie macierzy pikseli. **Diabetes** (*scikit-learn*) to dane regresyjne o cukrzycy, uwzględniające cechy demograficzne i biomedyczne. **Linnerud** (*scikit-learn*) obejmuje dane wielowymiarowe dotyczące parametrów fitnessowych. **Auto MPG** (Kaggle) zawiera dane regresyjne o zużyciu paliwa samochodów na podstawie ich parametrów [1]. **Auto Insurance in Sweden** (Kaggle) to dane regresyjne o kosztach ubezpieczeń samochodowych w Szwecji [2]. **Blood Transfusion Dataset** (Kaggle) opisuje dane klasyfikacyjne dotyczące skłonności dawców do oddania krwi [3]. Podczas analizy wstępnej wykryto znaczące ograniczenia w zbiorach **Linnerud**, **Auto MPG** które uniemożliwiły ich skuteczne wykorzystanie w ramach badania.

### 2.1 Przetwarzanie danych

Dane zostały znormalizowane przy użyciu *StandardScaler*, aby zapewnić, że wszystkie cechy mają średnią 0 i odchylenie standardowe 1, co zapobiega dominacji cech o dużej skali w procesie optymalizacji hiperparametrów. W przypadku zbiorów z Kaggle konieczne było dodatkowe wstępne przetwarzanie, takie jak usunięcie brakujących wartości i konwersja danych do formatu numerycznego.

## 3 Metodyka

### 3.1 Wybrane Algorytmy

Tabela 1 przedstawia zestawienie algorytmów uczenia maszynowego wykorzystanych w badaniu wraz z optymalizowanymi hiperparametrami i ich zakresami.

Tabela 1: Wybrane algorytmy i zakresy hiperparametrów

Algorytm	Hiperparametr	Zakres wartości
Random Forest	<code>n_estimators</code>	10, 50, 100, 200
	<code>max_depth</code>	5, 10, 20, 30
	<code>min_samples_split</code>	2, 5, 10
	<code>min_samples_leaf</code>	1, 2, 4
XGBoost	<code>n_estimators</code>	50, 100, 200
	<code>learning_rate</code>	0.01, 0.1, 0.2
	<code>max_depth</code>	3, 5, 7, 10
	<code>subsample</code>	0.6, 0.8, 1.0
ElasticNet	<code>alpha</code>	0.01, 0.1, 1.0
	<code>l1_ratio</code>	0.1, 0.5, 0.9
	<code>max_iter</code>	1000, 2000, 5000
Gradient Boosting	<code>n_estimators</code>	50, 100, 200
	<code>learning_rate</code>	0.01, 0.1, 0.2
	<code>max_depth</code>	3, 5, 7

### 3.2 Techniki Tunowania

W eksperymencie zastosowano trzy metody optymalizacji hiperparametrów:

**Grid Search** — Klasyczne przeszukiwanie siatki hiperparametrów, testujące wszystkie możliwe kombinacje zadanych wartości. Wykorzystano `GridSearchCV` z biblioteki `scikit-learn` z 3-krotną walidacją krzyżową.

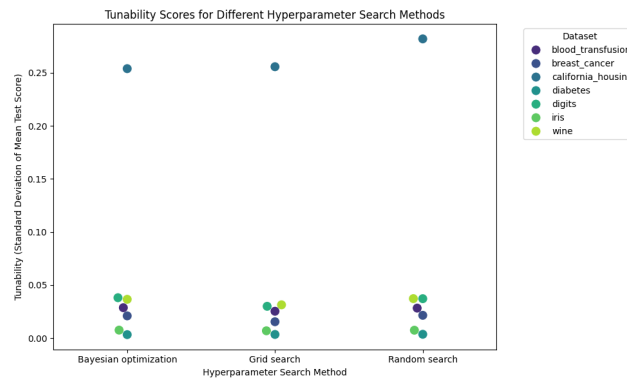
**Random Search** — Losowy wybór kombinacji hiperparametrów, pozwalający na szybsze przeszukiwanie większej przestrzeni parametrów. Użyto `RandomizedSearchCV` z `scikit-learn`.

**Optymalizacja Bayesowska** — Zaawansowana metoda wykorzystująca modele probabilistyczne do wyboru najbardziej obiecujących obszarów przestrzeni parametrów. Implementacja oparta na `BayesSearchCV` z `scikit-optimize` aktualizowała priorytety na podstawie wyników kolejnych iteracji.

## 4 Wyniki

Na rysunku 1 przedstawiono tunowalność algorytmów dla różnych metod tunowania hiperparametrów. Zbieżność wyników w zależności od liczby iteracji pokazano na rysunku 2, a różnice krytyczne pomiędzy metodami tunowania hiperparametrów przedstawiono na rysunku 3. Na rysunkach przedstawiono wyniki analiz, ilustrujące tunowalność algorytmów, porównanie metod tunowania, zbieżność wyników oraz różnice krytyczne między metodami.

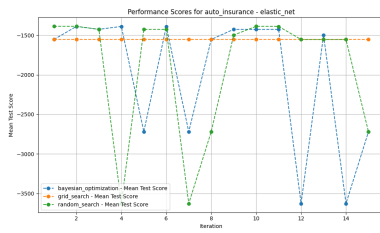
### 4.1 Tunowalność Algorytmów



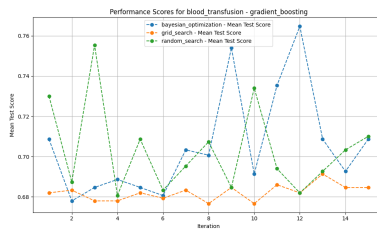
Rysunek 1: Wykres przedstawia tunowalność algorytmów mierzonych jako odchylenie standardowe wyników modelu na różnych zbiorach danych dla trzech metod tunowania hiperparametrów: Optymalizacji Bayesowskiej, Grid Search i Random Search. Każdy punkt reprezentuje wynik dla jednej kombinacji algorytmu i zbioru danych. Pozycja pionowa punktu pokazuje stopień tunowalności, gdzie wyższe wartości oznaczają większą wrażliwość algorytmu na zmiany hiperparametrów. Wyniki wskazują, że Optymalizacja Bayesowska charakteryzuje się bardziej przewidywalnymi wynikami niż Grid Search i Random Search, które mają większe rozproszenie.

XGBoost i Gradient Boosting wykazują wyższą tunowalność w porównaniu z ElasticNet. Tunowalność była największa dla Optymalizacji Bayesowskiej.

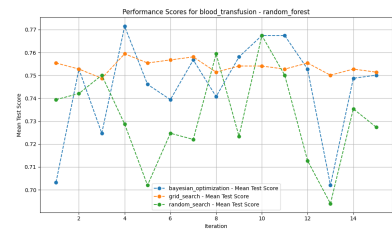
## 4.2 Zbieżność wyników i stabilizacja modeli



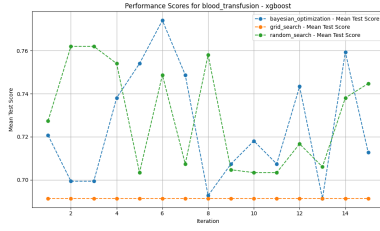
(a) Auto Insurance - ElasticNet



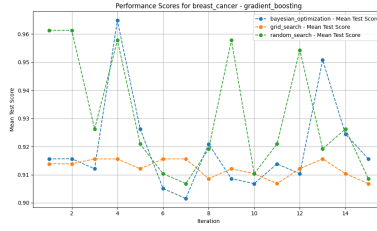
(b) Blood Transfusion - Gradient Boosting



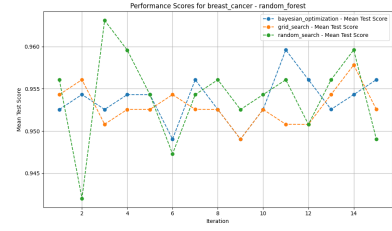
(c) Blood Transfusion - Random Forest



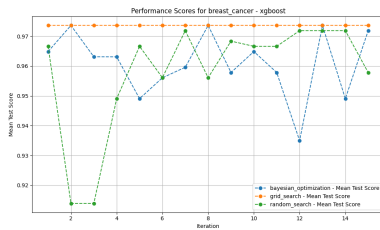
(d) Blood Transfusion - XGBoost



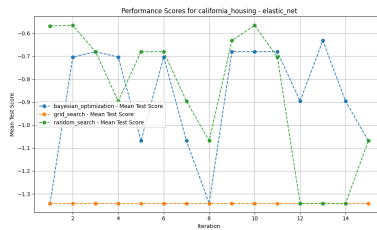
(e) Breast Cancer - Gradient Boosting



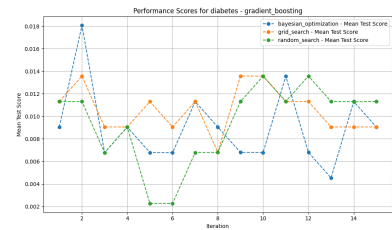
(f) Breast Cancer - Random Forest



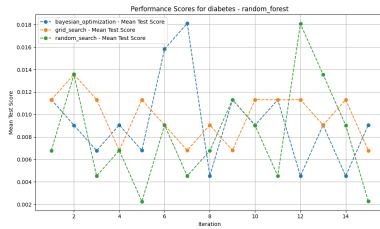
(g) Breast Cancer - XGBoost



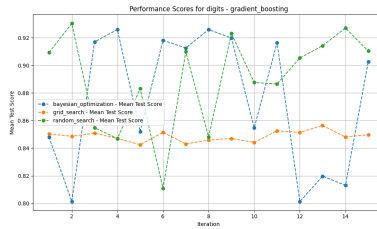
(h) California Housing - ElasticNet



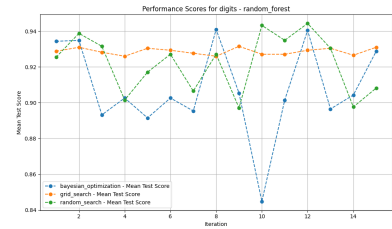
(i) Diabetes - Gradient Boosting



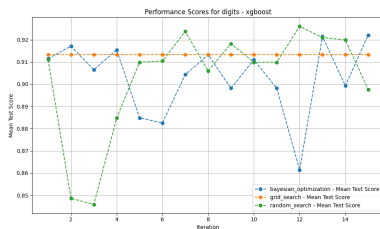
(j) Diabetes - Random Forest



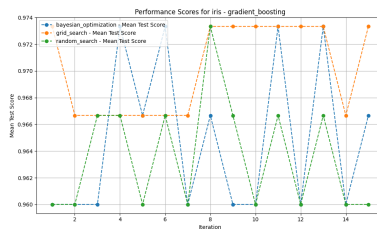
(k) Digits - Gradient Boosting



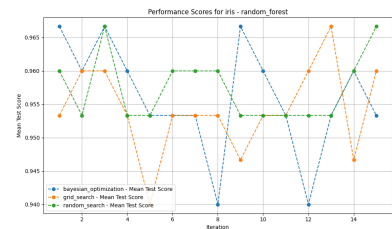
(l) Digits - Random Forest



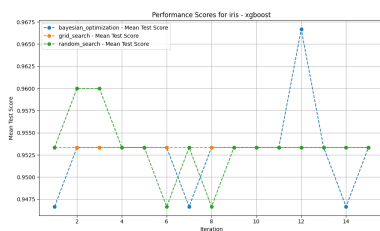
(m) Digits - XGBoost



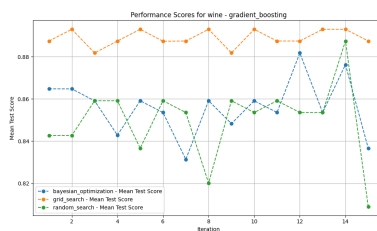
(n) Iris - Gradient Boosting



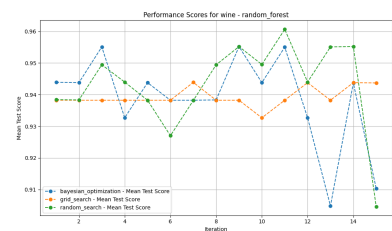
(o) Iris - Random Forest



(p) Iris - XGBoost



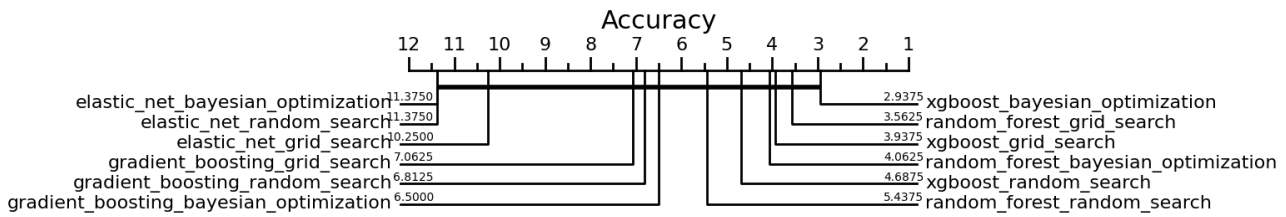
(q) Wine - Gradient Boosting



(r) Wine - Random Forest

Rysunek 2: Wykresy przedstawiające wyniki wydajności algorytmów na różnych zbiorach danych i metodach tunowania hiperparametrów.

### 4.3 Diagram Różnic Krytycznych



Rysunek 3: Diagram różnic krytycznych (CD-Diagram) przedstawia porównanie metod tunowania hiperparametrów dla różnych modeli na podstawie ich średnich rang. Oś pozioma reprezentuje średnie rangi poszczególnych metod, gdzie niższa wartość oznacza lepszą skuteczność. Linie poziome łączą metody, między którymi nie stwierdzono statystycznie istotnych różnic według testu Wilcozona z poprawką Holm-Bonferroni. Diagram został wygenerowany na podstawie wyników dla 8 zbiorów danych.

Diagram różnic krytycznych przedstawia analizę post-hoc przeprowadzoną na podstawie testu Friedmana, który odrzucił hipotezę zerową o braku różnic między metodami. Następnie zastosowano test Wilcozona z poprawką Holm-Bonferroni do analizy parowych różnic między metodami.

Grube linie poziome w diagramie wskazują grupy metod, pomiędzy którymi nie można odrzucić hipotezy zerowej na poziomie istotności  $\alpha = 0.05$ . Oznacza to, że nie stwierdzono statystycznie istotnych różnic w skuteczności między tymi metodami.

Z diagramu wynika, że:

- *xgboost\_grid\_search* oraz *xgboost\_bayesian\_optimization* osiągają jedno z najwyższych średnich rang, co sugeruje ich wysoką skuteczność.
- Metody związane z *ElasticNet* mają najniższe rangi, co oznacza relatywnie niższą skuteczność w porównaniu do innych metod.
- Grube linie łączące grupy metod (np. *xgboost\_grid\_search* i *random\_forest\_bayesian\_optimization*) wskazują, że różnice w ich wynikach nie są statystycznie istotne.

Podsumowując, przy poziomie istotności  $\alpha = 0.05$ , diagram ilustruje, że pomimo widocznych różnic w średnich rangach nie można formalnie stwierdzić, że metody połączone poziomymi liniami różnią się od siebie w sposób statystycznie istotny. Oznacza to, że choć niektóre metody osiągają wyraźnie wyższe średnie rangi (np. metody *xgboost*), różnice te mogą wynikać z przypadkowych fluktuacji danych.

## 5 Wnioski

- Zakresy hiperparametrów dla poszczególnych algorytmów, zostały dostosowane na podstawie literatury oraz wyników eksperymentalnych, co potwierdziło ich znaczenie w procesie optymalizacji.
- Algorytmy takie jak XGBoost i Gradient Boosting wykazały wyższą tunowalność, co wskazuje na ich większą podatność na dostosowanie parametrów, podczas gdy algorytmy takie jak ElasticNet wykazywały mniejszy wpływ zmian hiperparametrów na wyniki.
- Metody losowania punktów, takie jak Random Search i Grid Search, wprowadzały większe rozproszenie wyników w porównaniu do Optymalizacji Bayesowskiej. Stwierdzono, że technika samplingu może wprowadzać bias, szczególnie dla bardziej złożonych algorytmów, takich jak XGBoost.
  - Diagram różnic krytycznych potwierdza, że metody optymalizacji dla XGBoost (*grid search* i *bayesian optimization*) osiągają najlepsze średnie rangi, co wskazuje na ich wysoką skuteczność w strojeniu hiperparametrów.
  - W przypadku algorytmu ElasticNet różnice między metodami tunowania były mniej istotne, co sugeruje, że wpływ hiperparametrów na wydajność modelu jest w tym przypadku ograniczony.
  - Grube linie na diagramie wskazują grupy metod, które są statystycznie równoważne pod względem wyników, co oznacza brak znaczących różnic między np. *random\_forest\_bayesian\_optimization* a *elastic\_net\_random\_search*.
  - Wyniki sugerują, że bardziej zaawansowane metody tunowania, takie jak Optymalizacja Bayesowska, mogą znacząco poprawić wyniki dla złożonych modeli, takich jak XGBoost, w porównaniu do bardziej tradycyjnych metod, jak Grid Search.

## Literatura

- [1] UCI Machine Learning Repository. Auto mpg dataset, 2020.
- [2] Kaggle Community. Auto insurance in sweden, 2020.
- [3] Kaggle Community. Blood transfusion dataset, 2020.