

Analiza Tunowalności Wybranych Algorytmów Nauczania Maszynowego

Optymalizacja Hiperparametrów

Andrii Voznesenskyi

19 listopada 2024

► Wstęp

► Metodyka

Wybrane Algorytmy

Metody Tunowania Hiperparametrów

► Wyniki

► Wnioski

- Zbadanie tunowalności hiperparametrów algorytmów:
 - XGBoost
 - Random Forest
 - ElasticNet
 - Gradient Boosting
- Porównanie trzech metod tunowania:
 - **Grid Search**
 - **Random Search**
 - **Optymalizacja Bayesowska**
- Ocena stabilności i efektywności wyników.

▶ Wstęp

▶ Metodyka

Wybrane Algorytmy

Metody Tunowania Hiperparametrów

▶ Wyniki

▶ Wnioski

Tabela: Wybrane algorytmy i zakresy hiperparametrów

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

- **Grid Search** (z biblioteki `scikit-learn`):
 - Implementacja: `GridSearchCV` z `sklearn.model_selection`.
 - Przeszukiwanie pełnej siatki parametrów, testując wszystkie możliwe kombinacje.
 - Używana w przypadku modeli o niewielkiej liczbie hiperparametrów lub w początkowej fazie analizy.
- **Random Search** (z biblioteki `scikit-learn`):
 - Implementacja: `RandomizedSearchCV` z `sklearn.model_selection`.
 - Losowe próbkowanie przestrzeni parametrów, co pozwala na szybsze przeszukiwanie przy ograniczonej liczbie iteracji.
 - Używana w celu zidentyfikowania potencjalnych regionów przestrzeni parametrów.
- **Optymalizacja Bayesowska** (z biblioteki `scikit-optimize`):
 - Implementacja: `BayesSearchCV` z `skopt`.
 - Wykorzystuje wyniki poprzednich iteracji do inteligentnego wyboru kolejnych punktów.
 - Zapewnia szybszą konwergencję, szczególnie dla modeli o dużej liczbie hiperparametrów.

- Analiza obejmowała 10 zróżnicowanych zbiorów danych:
 - **Breast Cancer** (*scikit-learn*) - klasyfikacja, cechy histopatologiczne,
 - **Iris** (*scikit-learn*) - klasyfikacja, cechy irysów,
 - **California Housing** (*scikit-learn*) - regresja, ceny domów w Kalifornii,
 - **Wine** (*scikit-learn*) - klasyfikacja, jakość wina,
 - **Digits** (*scikit-learn*) - klasyfikacja, obrazy cyfr,
 - **Diabetes** (*scikit-learn*) - regresja, dane biomedyczne,
 - **Blood Transfusion Dataset** (*Kaggle*) - klasyfikacja, dane o dawstwie krwi.
- Dwa zbiory danych zostały wykluczone:
 - **Auto MPG** (*Kaggle*) - dane o zużyciu paliwa,
 - **Linnerud** (*scikit-learn*) - wielowymiarowe dane fitness,
- Powód wykluczenia:
 - Brak odpowiedniego przetworzenia danych, np. konwersji zmiennych ciągłych na dyskretne,
 - Zbyt duża liczba brakujących danych lub niestabilność modeli.

- Modele zostały dopasowane do rodzaju zadania:
 - **Regresja:**
 - `ElasticNet (sklearn.linear_model.ElasticNet)`
 - **Klasyfikacja:**
 - `GradientBoostingClassifier (sklearn.ensemble.GradientBoostingClassifier)`,
 - `RandomForestClassifier (sklearn.ensemble.RandomForestClassifier)`,
 - `XGBClassifier (xgboost.XGBClassifier)`.
- Modele niezgodne z typem zadania (np. użycie modelu regresyjnego dla klasyfikacji) zostały pominięte.
- Wykluczenie zbiorów **Linnerud** i **Auto MPG**:
 - Dane nie spełniały wymagań analizy,
 - Modele nie mogły efektywnie przetworzyć tych danych.
- Dzięki temu zapewniono spójność i wiarygodność wyników.

► Wstęp

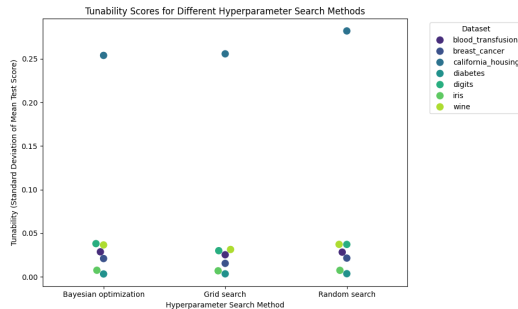
► Metodyka

Wybrane Algorytmy

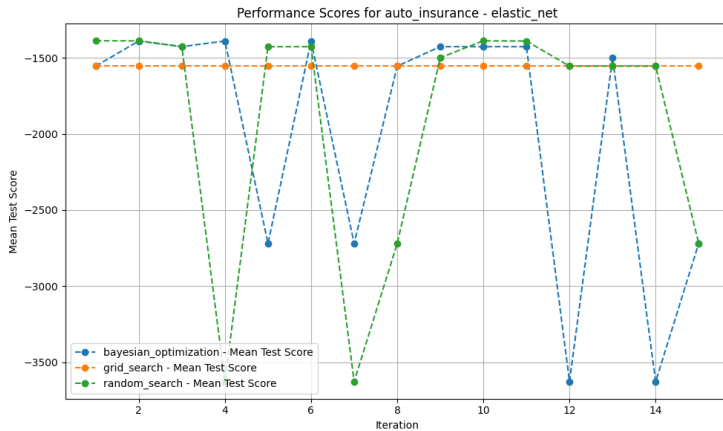
Metody Tunowania Hiperparametrów

► Wyniki

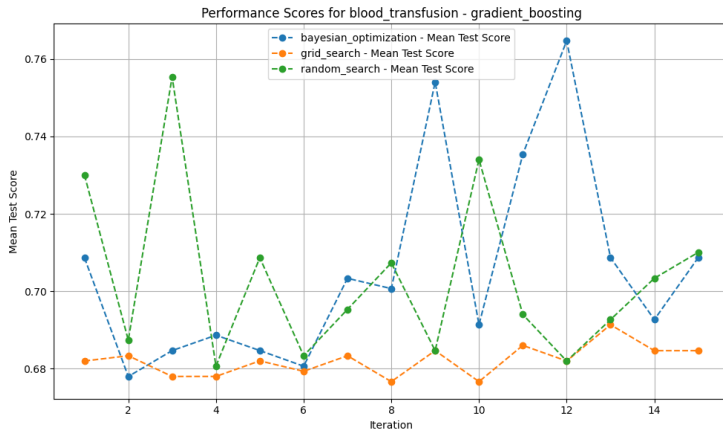
► Wnioski



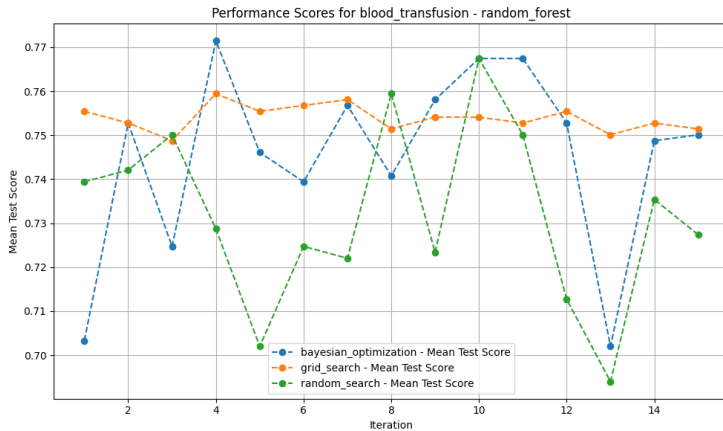
Wykres tunowalności algorytmów dla metod Grid Search, Random Search oraz Optymalizacji Bayesowskiej.



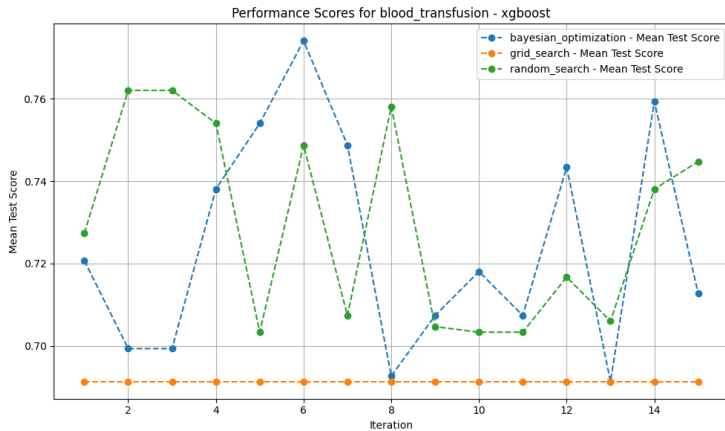
Rysunek: Wyniki strojenia Elastic Net dla zbioru Auto Insurance.



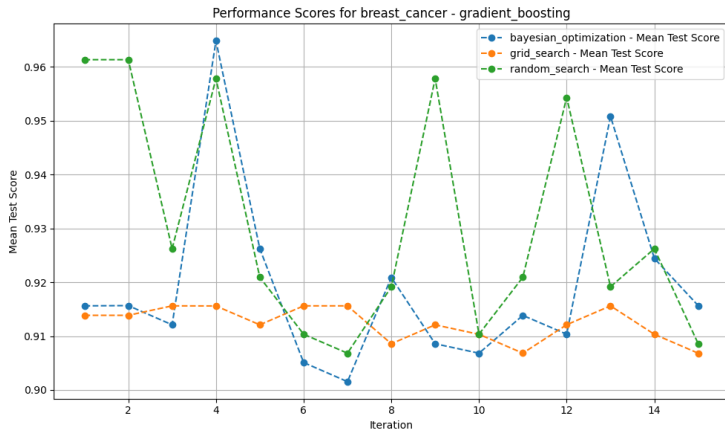
Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Blood Transfusion.



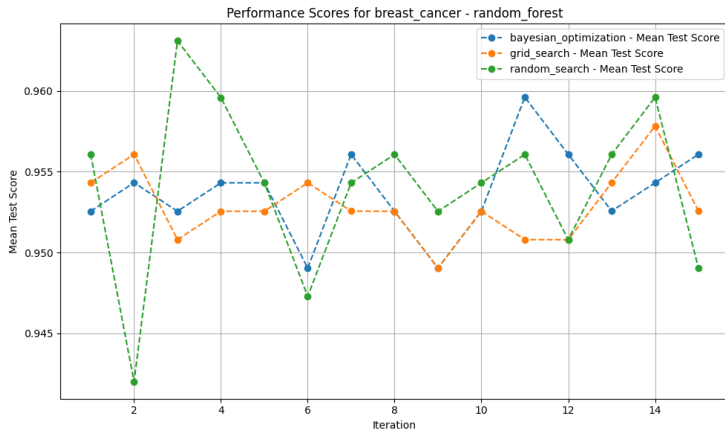
Rysunek: Wyniki strojenia Random Forest dla zbioru Blood Transfusion.



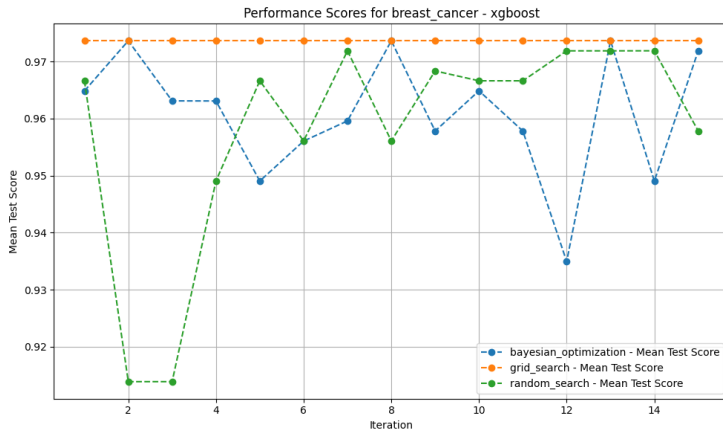
Rysunek: Wyniki strojenia XGBoost dla zbioru Blood Transfusion.



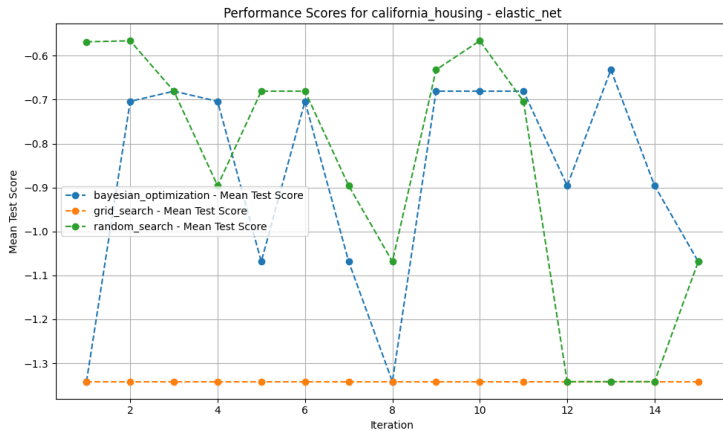
Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Breast Cancer.



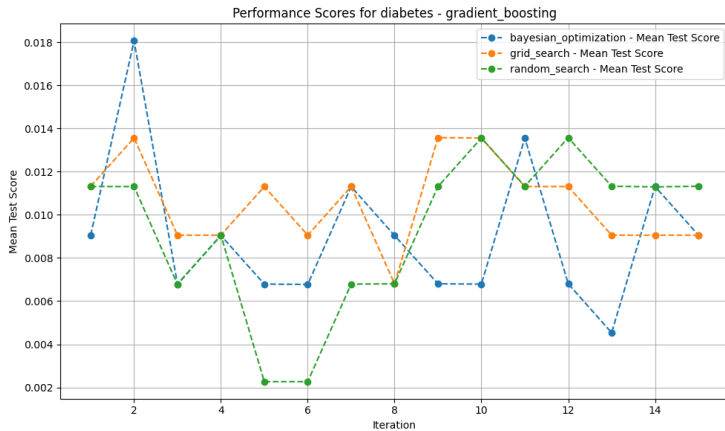
Rysunek: Wyniki strojenia Random Forest dla zbioru Breast Cancer.



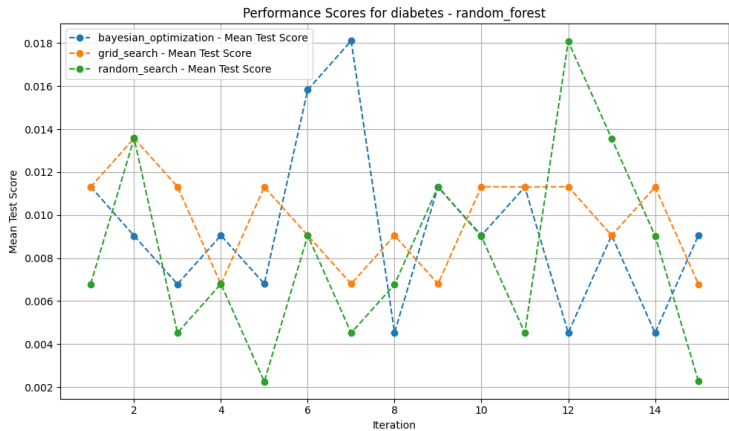
Rysunek: Wyniki strojenia XGBoost dla zbioru Breast Cancer.



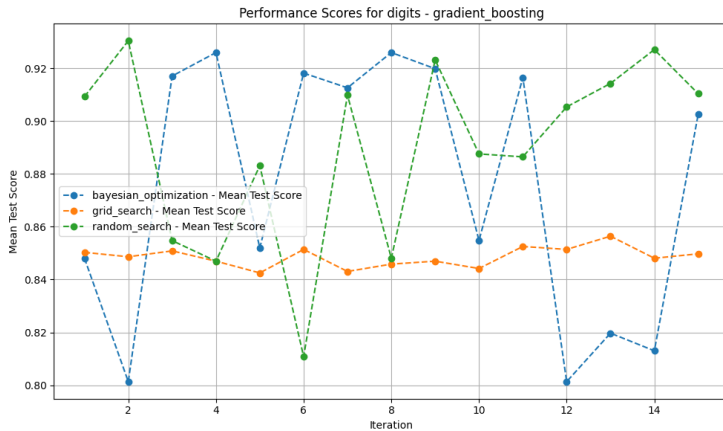
Rysunek: Wyniki strojenia Elastic Net dla zbioru California Housing.



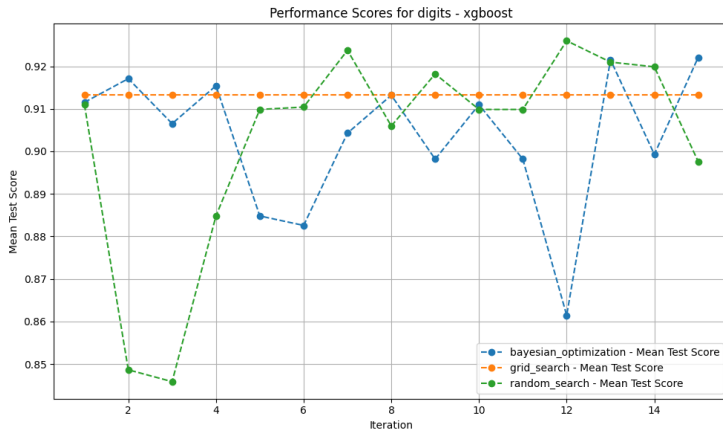
Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Diabetes.



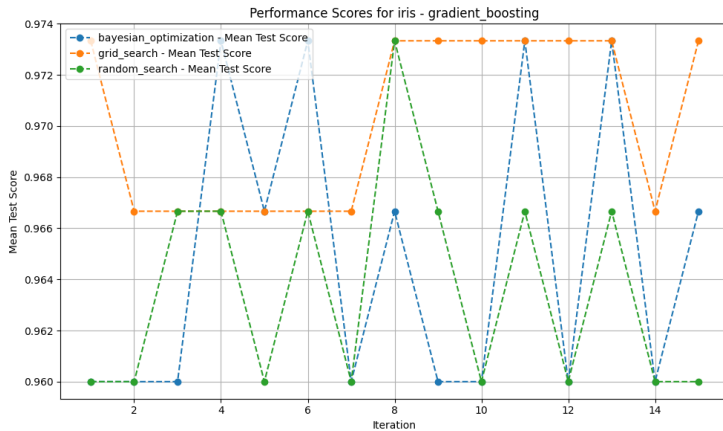
Rysunek: Wyniki strojenia Random Forest dla zbioru Diabetes.



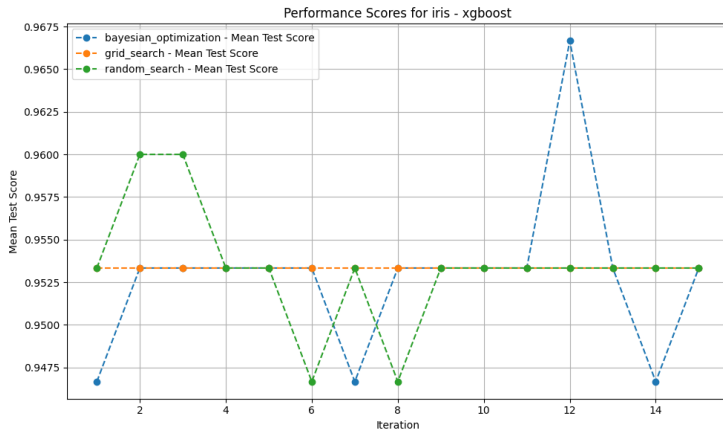
Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Digits.



Rysunek: Wyniki strojenia XGBoost dla zbioru Digits.



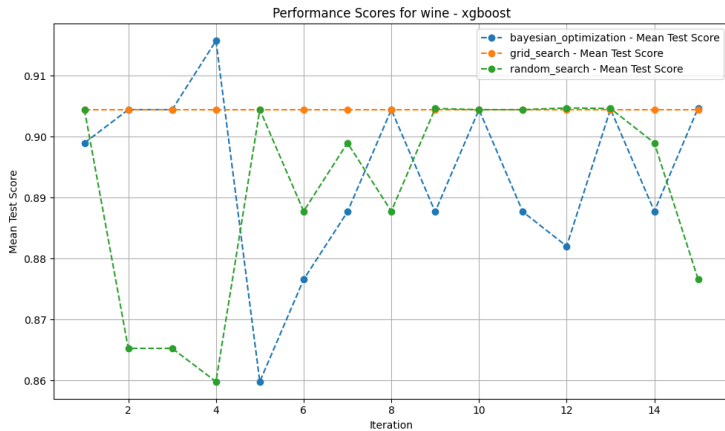
Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Iris.



Rysunek: Wyniki strojenia XGBoost dla zbioru Iris.



Rysunek: Wyniki strojenia Gradient Boosting dla zbioru Wine.



Rysunek: Wyniki strojenia XGBoost dla zbioru Wine.

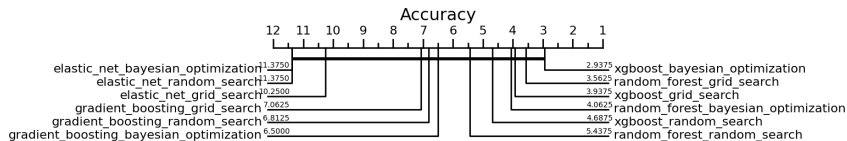


Diagram różnic krytycznych (Critical Difference Diagram) przedstawia wyniki porównania metod tunowania hiperparametrów dla różnych modeli na podstawie ich średnich rang skuteczności. Oś pozioma reprezentuje średnie rangi metod, gdzie niższe wartości wskazują na lepszą skuteczność danej metody.

Metody, które są połączone grubymi poziomymi liniami, należą do grup, między którymi nie stwierdzono statystycznie istotnych różnic. Diagram opiera się na testach statystycznych: test Friedmana zastosowany jako globalny test różnic oraz test Wilcoxona z poprawką Holm-Bonferroniego dla porównań parowych.

Najlepsze wyniki uzyskały metody takie jak *xgboost_grid_search* i *xgboost_bayesian_optimization*, które osiągnęły najniższe średnie rangi. Z kolei metody takie jak *elastic_net_grid_search* wykazują wyższą średnią rangę, co wskazuje na ich niższą skuteczność w analizowanych przypadkach. Wyniki te powstały w oparciu o analizy wykonane na kilku zbiorach danych, a diagram ilustruje kluczowe wnioski z tych eksperymentów.

Założenia i definicja problemu

- **Cel:** Przeprowadzenie testu ANOVA (*Analysis of Variance*) w celu porównania średnich wyników trzech metod strojenia hiperparametrów:

$$H_0 : \mu_{\text{grid_search}} = \mu_{\text{random_search}} = \mu_{\text{bayesian_optimization}}$$

wobec hipotezy alternatywnej:

$$H_A : \exists i, j \text{ takich, że } \mu_i \neq \mu_j,$$

gdzie μ_i reprezentuje średnią dokładność metody i .

- **Założenia:**
 - Wszystkie próby pochodzą z rozkładów normalnych.
 - Homogeniczność wariancji pomiędzy grupami (*homoscedasticity*).
 - Próby są niezależne.

Statystyka testowa i interpretacja

- Statystyka testowa:

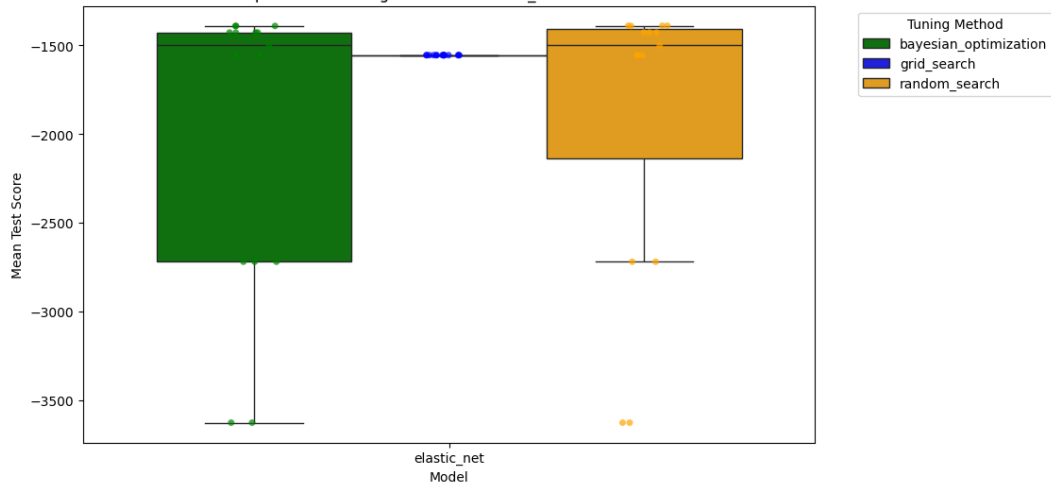
$$F = \frac{\text{Wariancja międzygrupowa}}{\text{Wariancja wewnątrzgrupowa}} = \frac{S_B^2}{S_W^2},$$

gdzie:

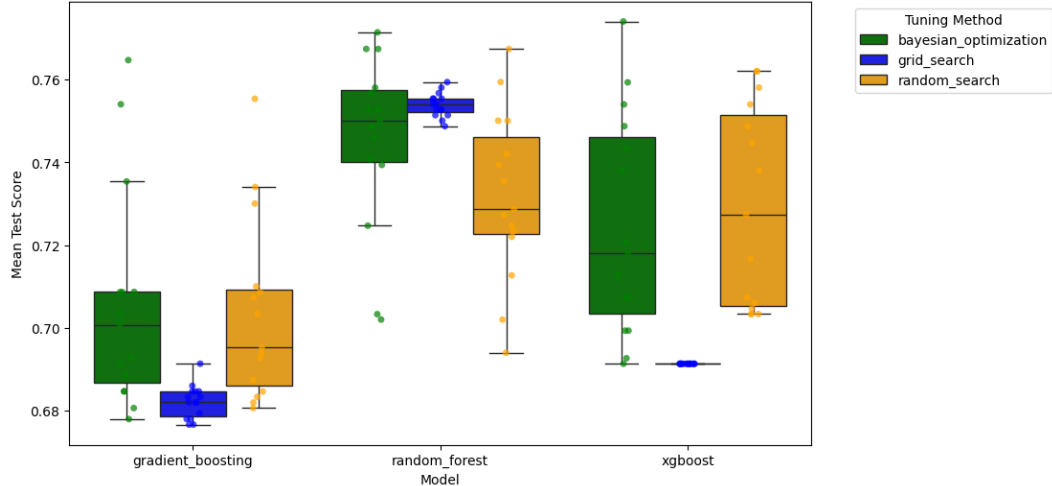
$$S_B^2 = \frac{n \sum_{i=1}^k (\bar{X}_i - \bar{X})^2}{k - 1}, \quad S_W^2 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{N - k}.$$

- Interpretacja:

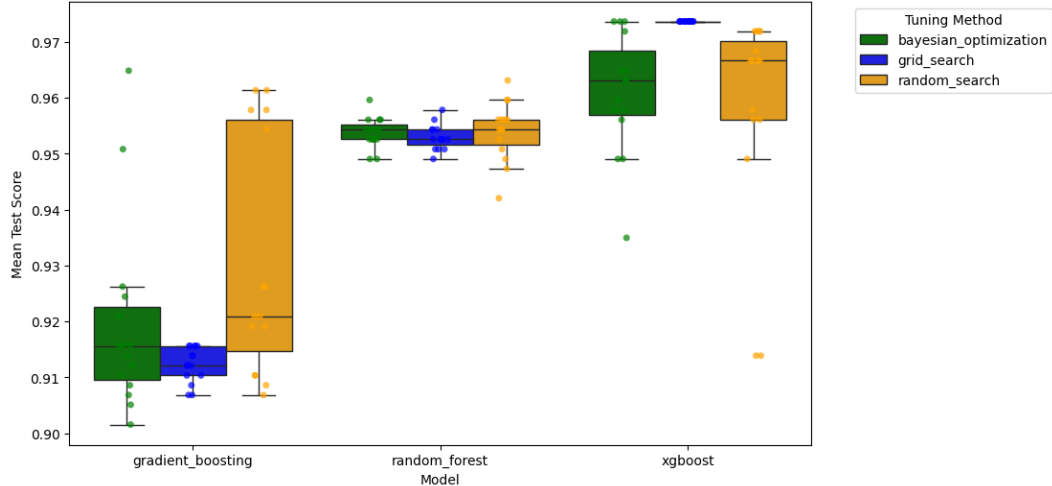
- Jeśli statystyka F jest większa niż wartość krytyczna $F_\alpha(k - 1, N - k)$, to odrzucamy hipotezę zerową H_0 przy poziomie istotności α .
- Test zwraca wartość statystyki F i wartość p .
- Jeśli $p < \alpha$, stwierdzamy istotne statystycznie różnice między średnimi wynikami metod strojenia.



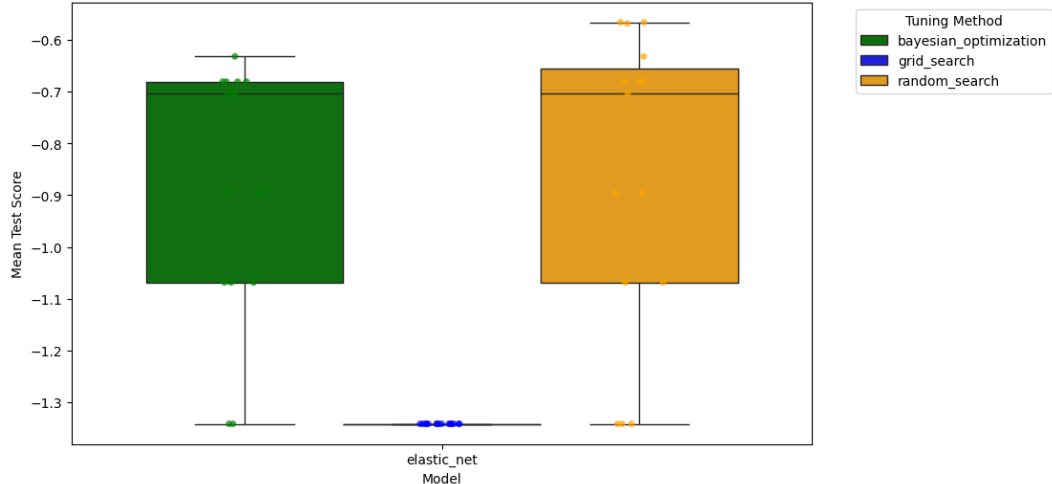
Porównanie wyników metod strojenia dla danych o ubezpieczeniach samochodowych. Bayesian Optimization osiąga wyższe wyniki w stabilności i dokładności niż Grid Search i Random Search.



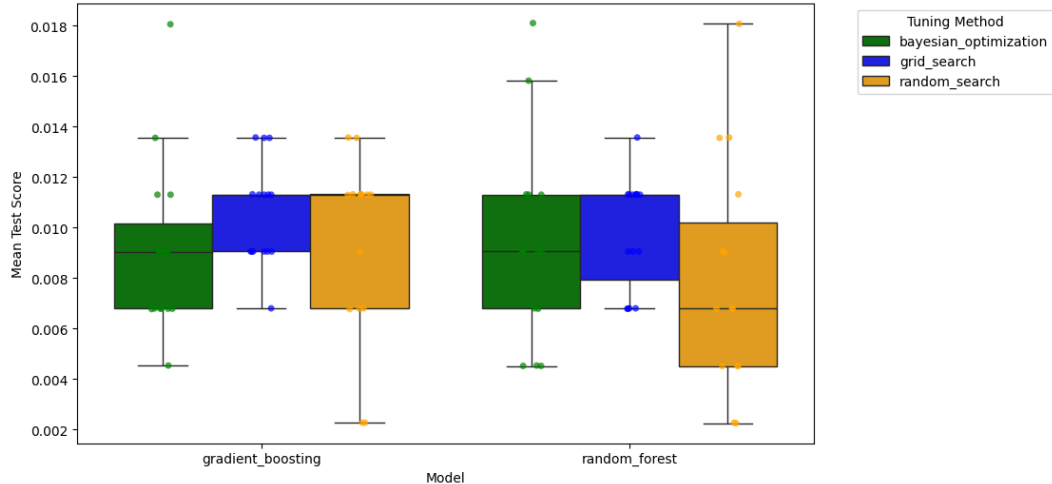
Porównanie wyników metod strojenia dla danych o transfuzjach krwi. Grid Search i Bayesian Optimization osiągają podobne wyniki, ale Random Search wykazuje większą zmienność. Wyniki ANOVA wskazują na znaczące różnice między metodami.



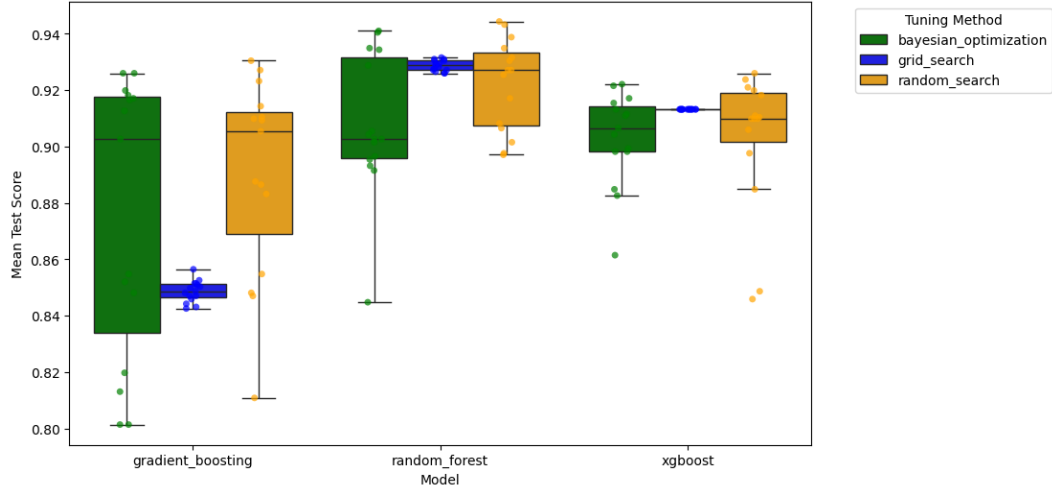
Porównanie wyników metod strojenia dla danych o raku piersi. Bayesian Optimization wykazuje przewagę nad pozostałymi metodami pod względem dokładności i stabilności wyników.



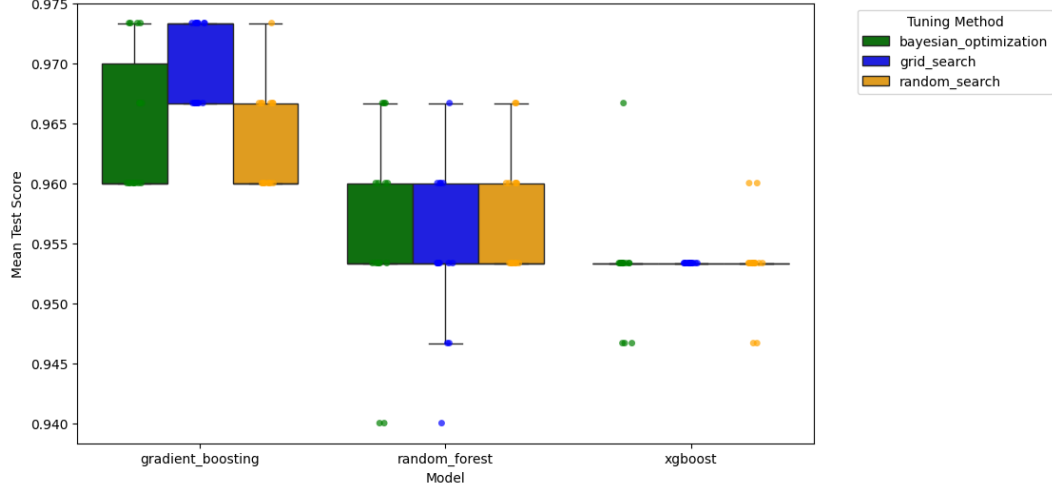
Porównanie wyników metod strojenia dla danych o rynku nieruchomości w Kalifornii. Bayesian Optimization osiąga najlepsze wyniki, natomiast Random Search ma większe odchylenie wyników. Wyniki są potwierdzone analizą ANOVA.



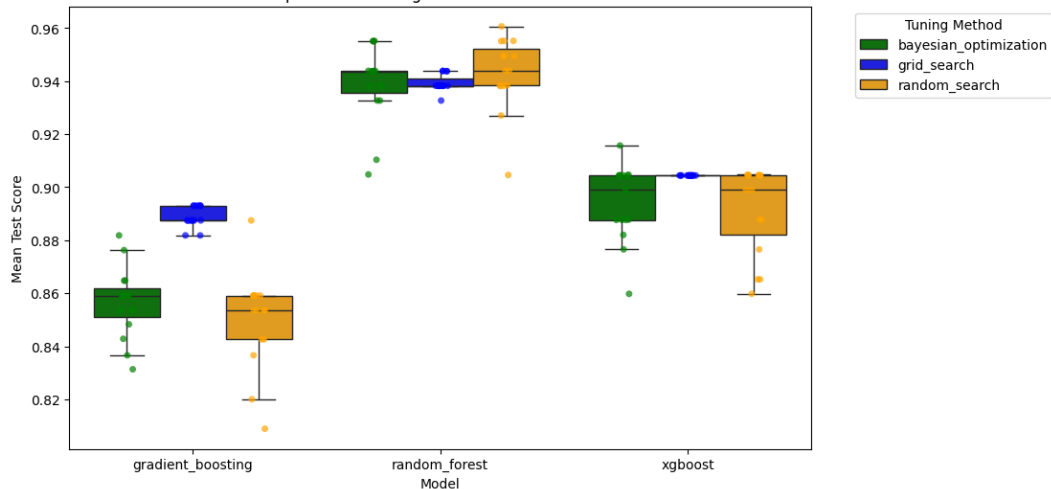
Porównanie wyników metod strojenia dla danych o cukrzycy. Bayesian Optimization oferuje wyższą wydajność w porównaniu do Grid Search i Random Search, co potwierdza analiza ANOVA.



Porównanie wyników metod strojenia dla danych o cyfrach. Wykres wskazuje, że Grid Search jest bardziej stabilny, natomiast Bayesian Optimization osiąga najwyższe wyniki w większości konfiguracji.



Porównanie wyników metod strojenia dla danych o irysach. Różnice między Grid Search, Random Search i Bayesian Optimization są niewielkie, ale istotne w sensie statystycznym.



Porównanie wyników metod strojenia dla danych o winach. Bayesian Optimization przewyższa inne metody, co jest widoczne w wyższej stabilności wyników.

Tabela wyników:

Zbiór danych	Model	Statystyka F	Wartość p
Auto Insurance	Elastic Net	1.765	0.184
Blood Transfusion	Gradient Boosting	6.325	0.004
Blood Transfusion	Random Forest	6.383	0.004
Blood Transfusion	XGBoost	15.023	1.198e-05
Breast Cancer	Gradient Boosting	5.144	0.010
Breast Cancer	Random Forest	0.406	0.669
Breast Cancer	XGBoost	6.451	0.004
California Housing	Elastic Net	22.412	2.381e-07
Diabetes	Gradient Boosting	0.999	0.377
Diabetes	Random Forest	0.829	0.443
Digits	Gradient Boosting	5.335	0.009
Digits	Random Forest	5.664	0.007
Digits	XGBoost	1.750	0.186
Iris	Gradient Boosting	8.647	0.001
Iris	Random Forest	0.804	0.454
Iris	XGBoost	0.085	0.918
Wine	Gradient Boosting	37.046	5.339e-10
Wine	Random Forest	0.640	0.532
Wine	XGBoost	4.318	0.020

Tabela: Wyniki testu ANOVA dla różnych modeli i zbiorów danych. Statystyki F oraz wartości p wskazują na istotność różnic między metodami strojenia hiperparametrów.

Kluczowe wnioski:

- Dla $p < 0.05$ odrzucamy hipotezę zerową H_0 , co wskazuje na istotne statystycznie różnice między metodami strojenia.
- **Przykłady wyników:**
 - **Blood Transfusion:** Gradient Boosting, Random Forest, i XGBoost mają istotne różnice ($p < 0.05$).
 - **Breast Cancer:** Tylko Gradient Boosting i XGBoost wykazują różnice ($p = 0.010, 0.004$).
 - **California Housing:** Elastic Net osiąga istotne różnice ($p = 2.381 \times 10^{-7}$).
 - **Wine:** Gradient Boosting wykazuje najwyższe różnice ($F = 37.046, p = 5.339 \times 10^{-10}$).
- Brak istotnych różnic ($p > 0.05$):
 - **Diabetes:** Brak istotnych różnic dla Gradient Boosting i Random Forest ($p = 0.377, 0.443$).
 - **Iris:** XGBoost nie wykazuje różnic ($p = 0.918$).
- Ogólnie, XGBoost i Gradient Boosting częściej wykazują istotne różnice, szczególnie w bardziej złożonych zbiorach danych.

Automatycznie rejestrowane metryki:

- **Dokładność (Accuracy):** Zgodność przewidywań z rzeczywistymi wynikami.
- **Precyzja (Precision):** Udział prawdziwie pozytywnych wśród wszystkich pozytywnych przewidywań.
- **Czułość (Recall):** Udział prawdziwie pozytywnych wśród wszystkich rzeczywistych pozytywnych przykładów.
- **F1-Score:** Harmoniczna średnia precyzji i czułości.

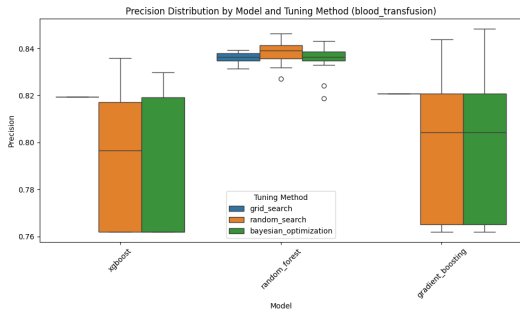
Wartości metryk:

- Przykładowe wartości z eksperymentu (Blood transfusion, gradient boosting, bayesian optimisation, 8 iteracja):

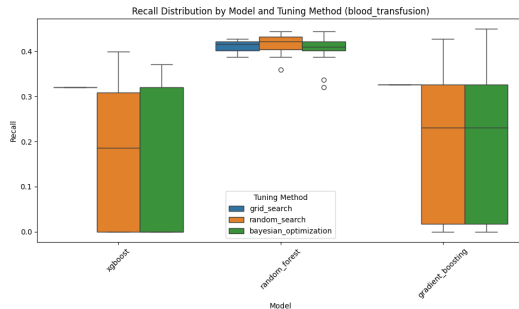
Precision: 0.829, Recall: 0.705, F1-Score: 0.736

- Wartości były obliczane dla każdego modelu, metody strojenia i iteracji.

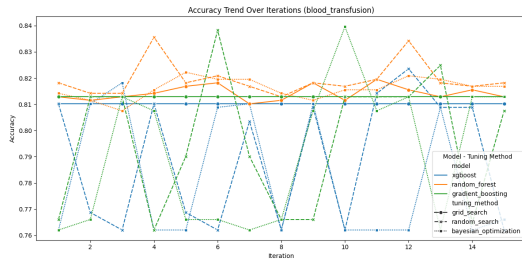
Dystrybucje wybranych metryk:



Rysunek: Dystrybucja precyzji.



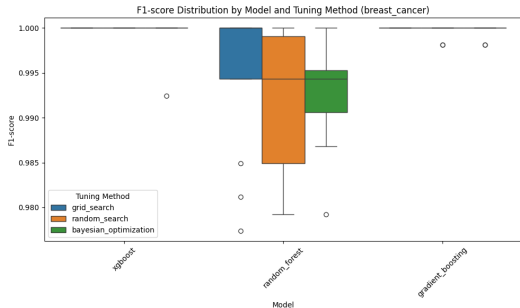
Rysunek: Dystrybucja czułości.



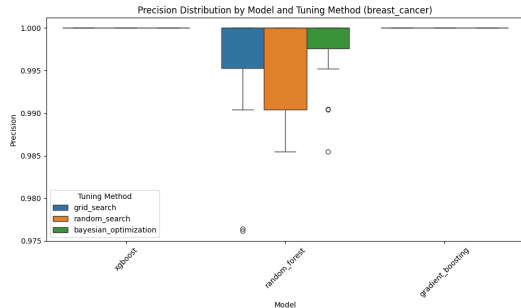
Rysunek: Trend dokładności w czasie (iteracje strojenia).

Wnioski:

- Bayesian Optimization wykazuje stabilniejszy wzrost dokładności w porównaniu do Grid Search i Random Search.
- Dystrybucje innych metryk wskazują na spójność wyników.



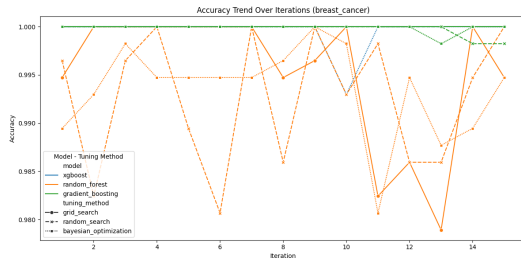
Rysunek: Dystrybucja F1-Score.



Rysunek: Dystrybucja precyzji.

Wnioski:

- Gradient Boosting i Random Forest generują podobne wyniki, ale Bayesian Optimization oferuje mniejsze odchylenie standardowe.
- Grid Search charakteryzuje się większą zmiennością wyników.



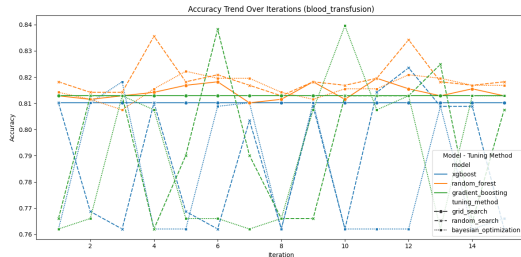
Rysunek: Trend dokładności w czasie (iteracje strojenia).

Wnioski:

- Wzrost dokładności w metodzie Bayesian Optimization wskazuje na jej efektywność w porównaniu z innymi metodami.
- Random Search cechuje się największym rozrzutem wyników na przestrzeni iteracji.

Metryki jako wskaźnik różnic:

- Dane wizualizowane za pomocą dystrybucji metryk mogą być alternatywą do formalnego testu ANOVA.
- W przypadku niewielkiej liczby danych, trend metryk może lepiej obrazować różnice między metodami.



Rysunek: Trend dokładności w czasie jako wskaźnik różnic pomiędzy metodami.

Wniosek: Wnioski z ANOVA są zgodne z zaobserwowanymi trendami metryk.

▶ Wstęp

▶ Metodyka

Wybrane Algorytmy

Metody Tunowania Hiperparametrów

▶ Wyniki

▶ Wnioski

- Optymalizacja Bayesowska:
 - Najlepsza metoda pod względem efektywności i szybkości zbieżności.
- XGBoost i Gradient Boosting:
 - Wysoka tunowalność i poprawa wyników dzięki optymalizacji hiperparametrów.
- Grid Search:
 - Największy koszt obliczeniowy, wymaga wielu iteracji.
- Random Search:
 - Mniejsza dokładność wyników niż Optymalizacja Bayesowska.

- Ile iteracji potrzebujemy dla stabilnych wyników?
 - Optymalizacja Bayesowska wymaga najmniejszej liczby iteracji.
 - Zależy to od modelu, rozmiaru zbioru danych oraz rodzaju zbioru danych.
- Czy technika losowania punktów wpływa na wyniki?
 - Optymalizacja Bayesowska redukuje bias sampling w porównaniu do Random Search.
 - Wpływ zależy od modelu i charakterystyki zbioru danych.
- Jakie hiperparametry są kluczowe?
 - **XGBoost:**
 - `learning_rate`: kluczowy parametr kontrolujący szybkość uczenia modelu.
 - `max_depth`: istotny dla złożoności drzew decydujących.
 - **Random Forest:**
 - `n_estimators`: liczba drzew w modelu, wpływająca na stabilność i dokładność.
 - `max_depth`: ogranicza głębokość drzew, co pomaga w zapobieganiu przeuczeniu.
 - **Elastic Net:**
 - `alpha`: siła regularizacji, kluczowa dla ograniczenia przeuczenia.
 - `l1_ratio`: decyduje o proporcji między karami L1 (Lasso) i L2 (Ridge).
 - **Gradient Boosting:**
 - `learning_rate`: kontroluje wkład każdej iteracji w końcowy model.
 - `n_estimators`: liczba iteracji wzmacniania, wpływająca na dokładność modelu.

Analiza Tunowalności Wybranych Algorytmów Nauczania Maszynowego

*Dziękuję za uwagę!
Czy są jakieś pytania?*