



# AUTOML

TUNOWALNOŚĆ ALGORYTMÓW I  
HIPERPARAMETRÓW





X G B O O S T

# SIATKA STARTOWA

Tabela 3: Siatka hiperparametrów modelu

Hiperparametr	Zakres wartości
model__n_estimators	randint(100, 1000)
model__learning_rate	uniform(0.01, 0.2)
model__max_depth	randint(1, 10)
model__subsample	uniform(0.1, 0.9)
model__min_child_weight	uniform(0, 7)
model__colsample_bytree	uniform(0.1, 0.9)
model__colsample_bylevel	uniform(0.1, 0.9)



# TUNOWALNOŚĆ ALGORYTMU

Dataset	Metoda Optymalizacji	
	Random Search	Bayes Search
diabeties (37)	0.01	0.023
creditg (31)	0.0071	-0.0083
spambase (44)	0.0016	0.0014
yeast (40597)	0.0013	-0.0074

Tabela 4: Tabela tunowalności dla *XGBoost*.



# TUNOWALNOŚĆ HIPERPARAMETRÓW

Dataset	n_estimators	learning_rate	max_depth	subsample	min_child_weight	colsample_bytree	colsample_bylevel
diabetes (37)	0.0071	0.0017	0.0076	0.0101	0.0084	0.0028	0.0021
creditg (31)	-0.0013	0.0015	0.0022	0.0062	0.0065	0.0017	0.0008
spambase (44)	0.0010	0.0007	0.0007	0.0012	0.0017	0.0007	0.0007
yeast (40597)	-0.0237	-0.0276	-0.0227	-0.0217	-0.0195	-0.0188	-0.0187

Tabela 5: Tunowalność hiperparametrów dla metody *Random Search* dla *XGBoost*.

Dataset	n_estimators	learning_rate	max_depth	subsample	min_child_weight	colsample_bytree	colsample_bylevel
diabetes (37)	-0.0003	-0.0043	0.0019	0.0049	0.0037	-0.0068	-0.0070
creditg (31)	-0.0267	-0.0305	-0.0222	-0.0261	-0.0187	-0.0194	-0.0224
spambase (44)	0.0016	0.0016	0.0015	0.0021	0.0028	0.0017	0.0015
yeast (40597)	-0.0238	-0.0243	-0.0027	-0.0233	-0.0249	-0.0220	-0.0231

Tabela 6: Tunowalność hiperparametrów dla metody *Bayes Search* dla *XGBoost*.



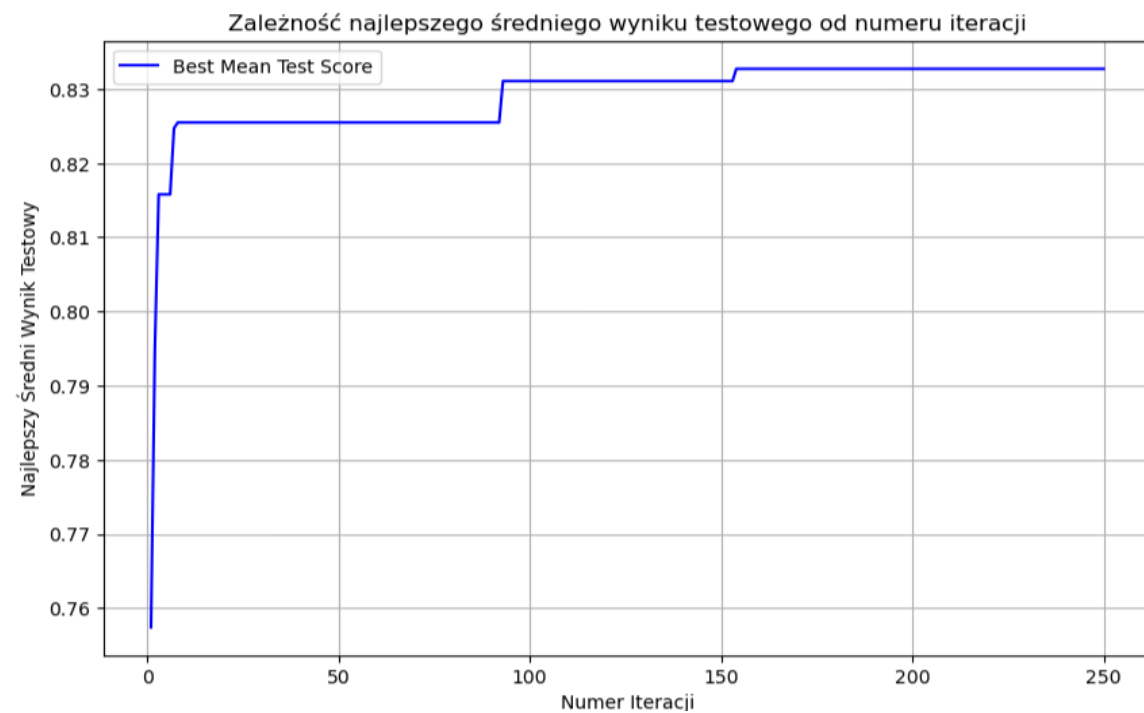
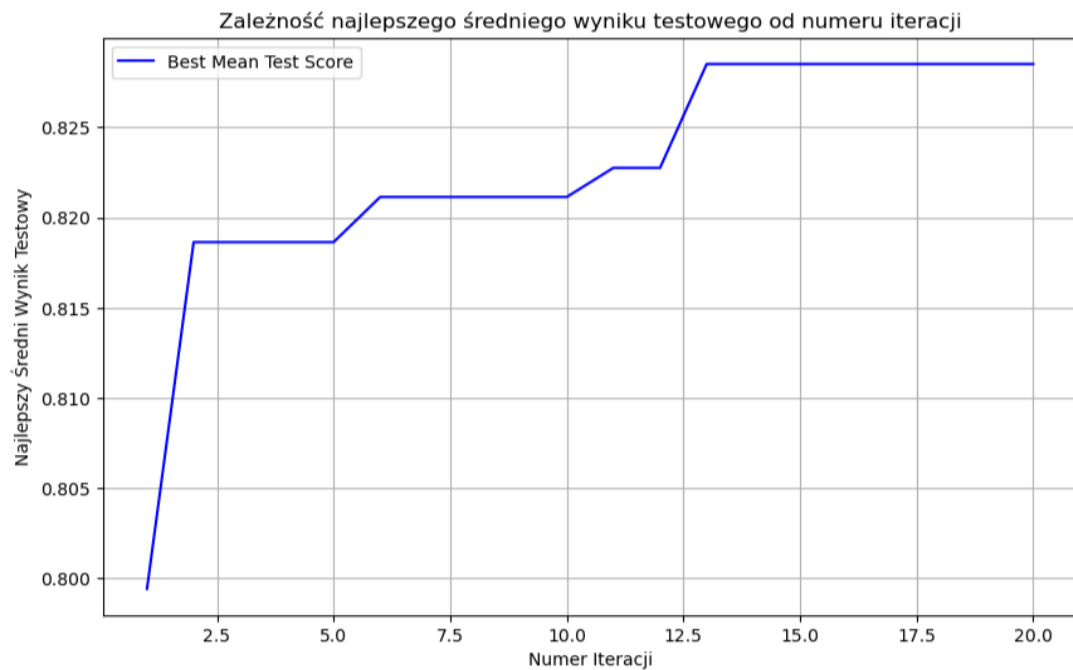
# KWANTYLE I WARTOŚCI OPTYMALNE

Hiperparametr	Zakres wartości	Kwantyl 5%	Kwantyl 95%	Optymalna wartość
model__n_estimators	randint(100, 1000)	103	960	809
model__learning_rate	uniform(0.01, 0.2)	0.010007	0.191541	0.010104
model__max_depth	randint(1, 10)	1.450000	10.0	9
model__subsample	uniform(0.1, 0.9)	0.100127	0.899934	0.559673
model__min_child_weight	uniform(0, 7)	0.025412	6.998327	2.133469
model__colsample_bytree	uniform(0.1, 0.9)	0.102904	0.875444	0.682921
model__colsample_bylevel	uniform(0.1, 0.9)	0.102904	0.864382	0.344919

Tabela 7: Siatka hiperparametrów modelu z zakresami kwantyli i optymalnymi wartościami.



# ZBIEŻNOŚĆ







# ELASTIC NET

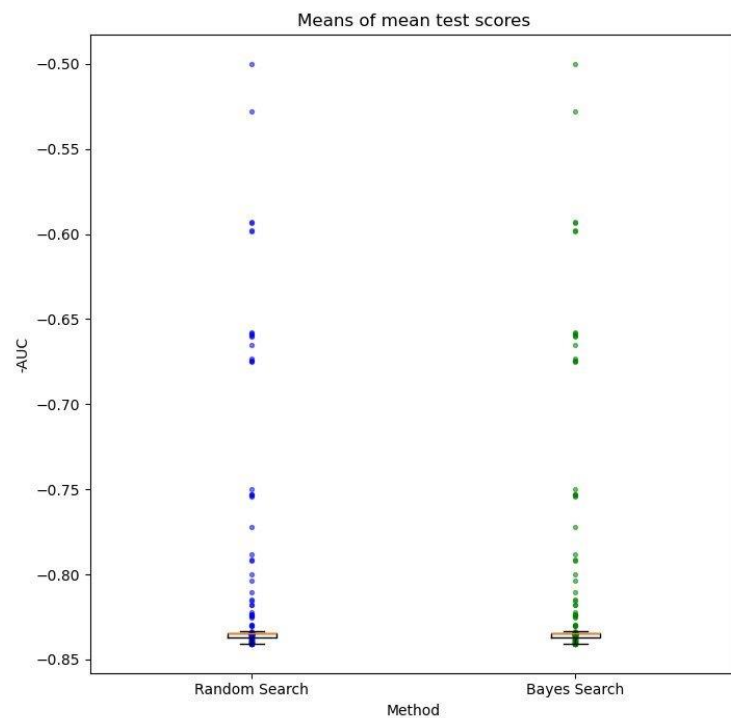
# ZAKRES POSZUKIWAŃ

Model Elastic Net posiada dwa główne hiperparametry:

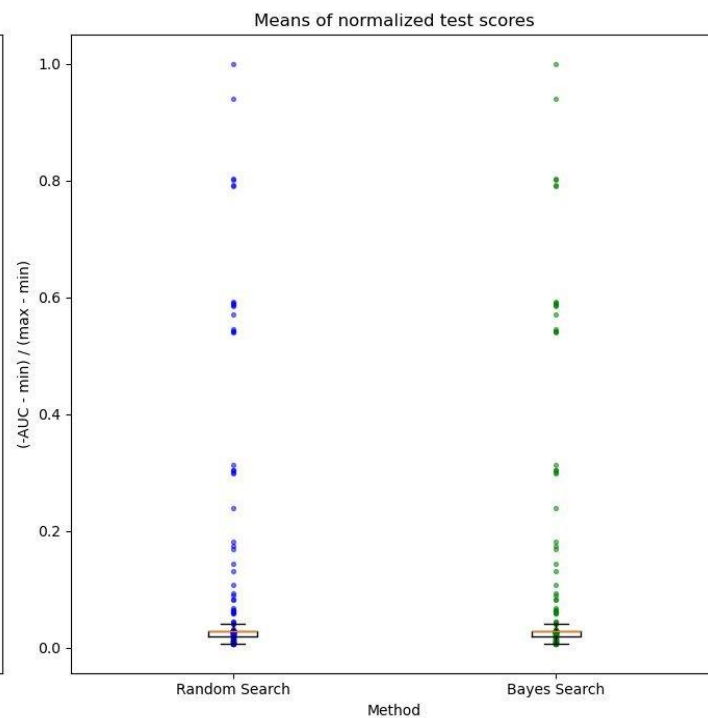
- **C** - rozkład logarytmiczno-jednostajny od  $2^{-10}$  do  $2^{10}$ ,
- **l1\_ratio** - rozkład jednostajny od 0 do 1.

Ich zakresy zaczerpnięto z artykułu *Tunability: Importance of hyperparameters of machine learning algorithms* (Probst, Bischl, and Boulesteix, 2018).

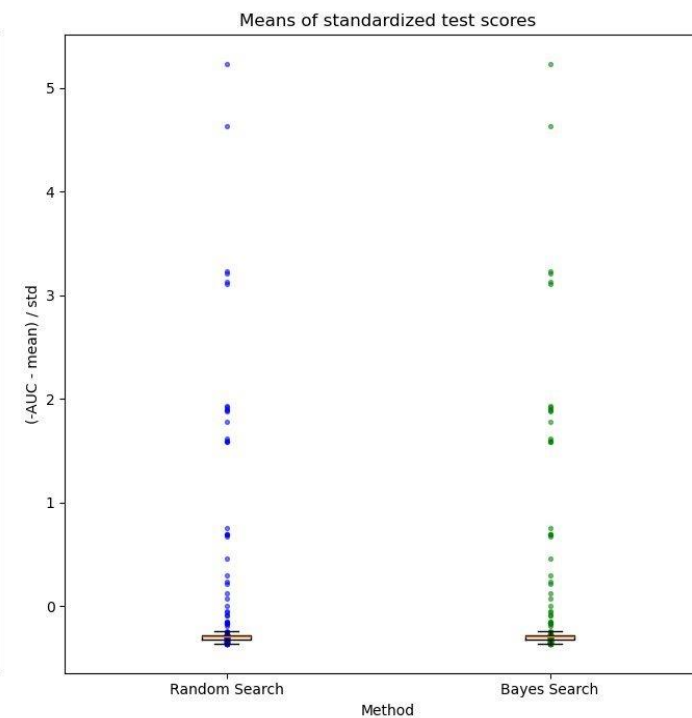
# WYBÓR OPTIMUM



średnie -AUC dla zestawu



średnie -AUC dla zestawu,  
znormalizowane wcześniej do przedziału  
[0,1] na każdym zbiorze danych



średnie -AUC dla zestawu,  
ustandaryzowane wcześniej na  
każdym zbiorze



# WYNIKI

Szukane	$\text{mean}(d_{RS})$	$\text{med}(d_{RS})$	$\text{mean}(d_{BO})$	$\text{med}(d_{BO})$	$\text{mean}(t_{it})$	$\text{med}(t_{it})$
C	0.000018	0.000041	-0.000875	0.000041	8.903920	6.951483
11	0.001472	0.002087	0.001693	0.002268	9.496289	8.589719
C i 11	0.001721	0.001759	0.002947	0.001435	11.054320	8.693008

$$t_{it} = \frac{t_{BO}}{t_{RS}} \cdot \frac{n_{RS}}{n_{BO}}$$

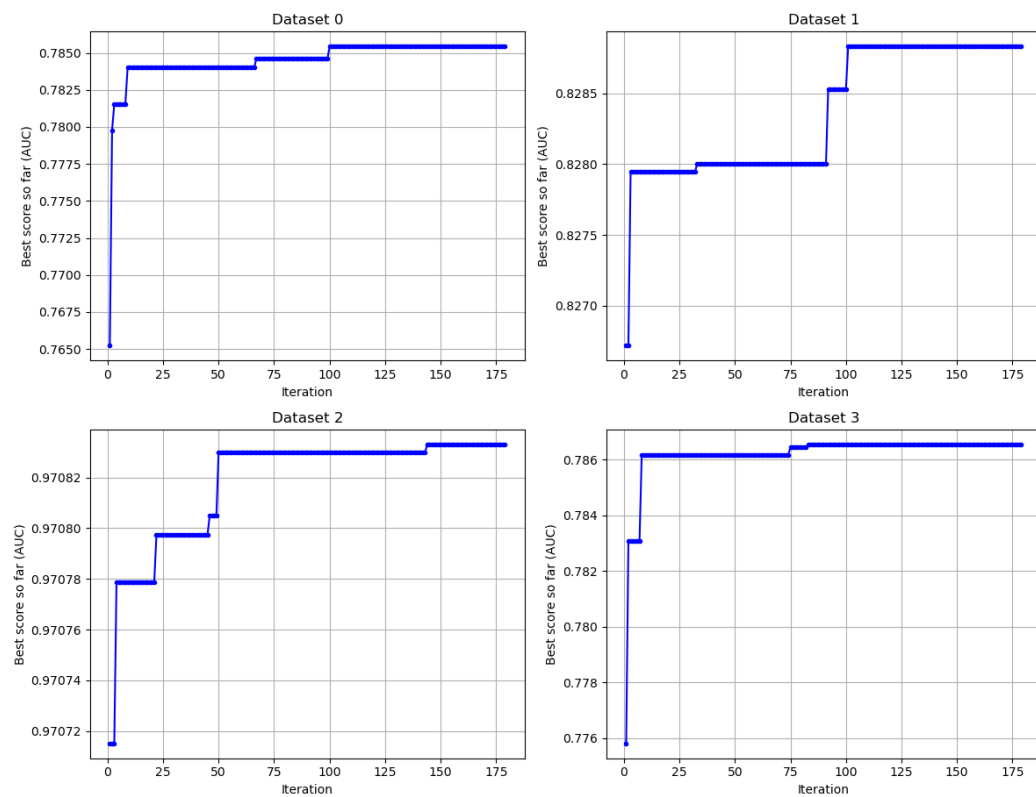
Średnie oraz mediany tunowalności hiperparametrów i algorytmu (szukane) oraz stosunku czasu iteracji metod.

Hiperparametr	$\text{mean}(d_{RS}^{rel})$	$\text{mean}(d_{BO}^{rel})$	$\text{med}(d_{RS}^{rel})$	$\text{med}(d_{BO}^{rel})$
C	0.010407	-0.296815	0.023350	0.028827
11	0.855382	0.574536	1.186648	1.579783

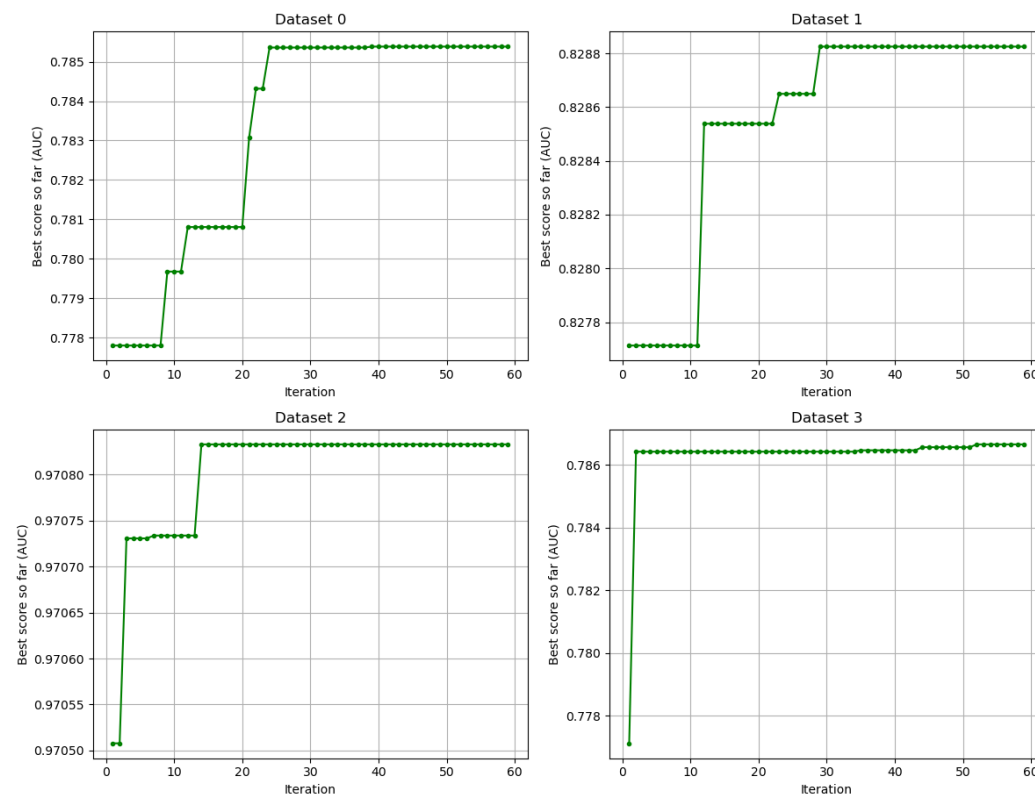
Średnie oraz mediany względnej tunowalności hiperparametrów.

# ZBIEŻNOŚĆ

Convergence Plots of Random Search

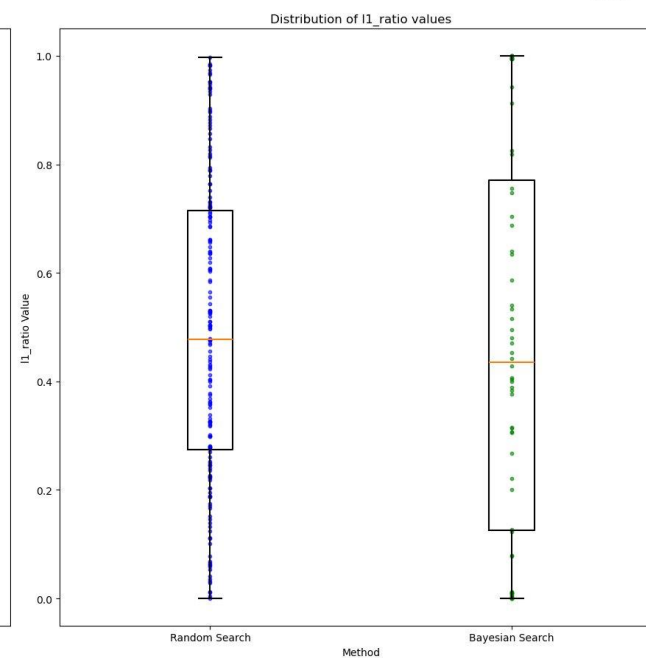
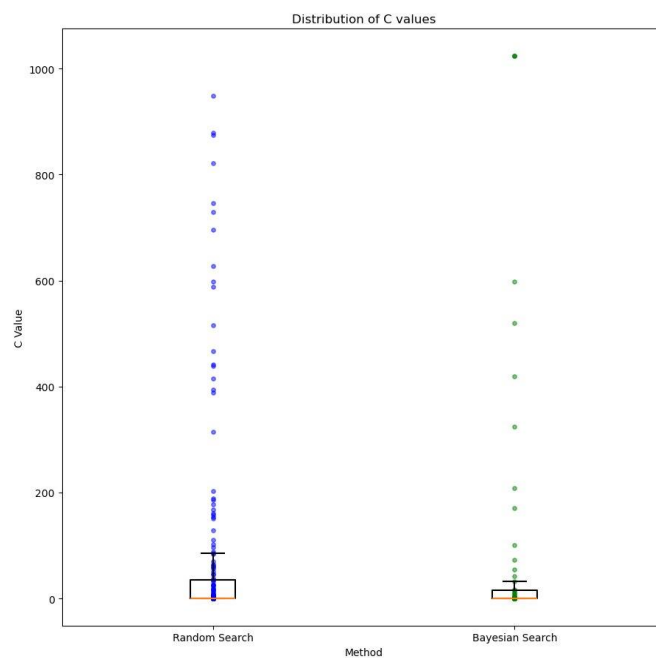
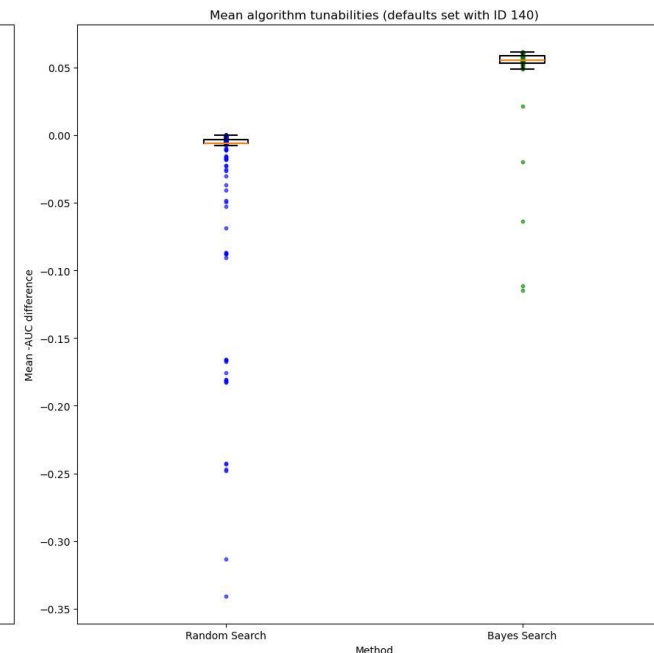
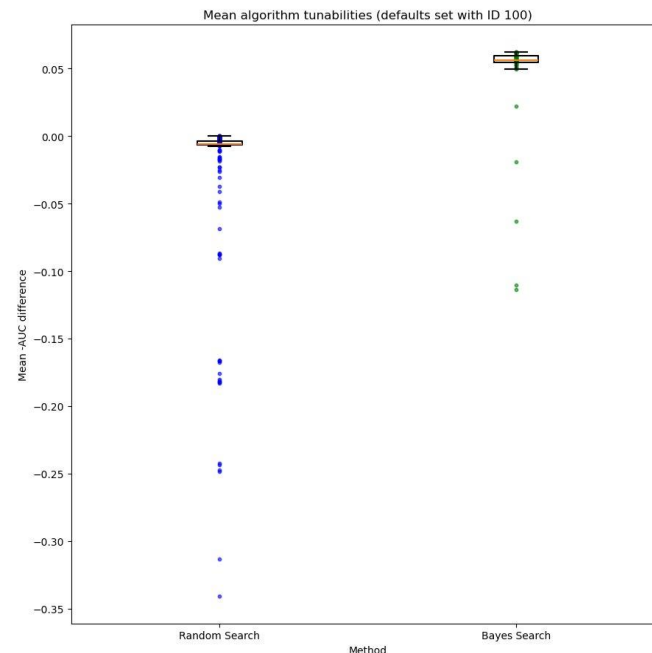


Convergence Plots of Bayesian Search



# ROZKŁADY

Średnia tunowalność algorytmu obliczona dla zestawów hiperparametrów przeszukiwanych podczas jego tunowania.



Dystrybucja przeszukiwanych wartości hiperparametrów podczas tunowania algorytmu dla każdej z metod samplingu.







# R A N D O M F O R E S T

# SIATKA STARTOWA

Tabela 4: Siatka hiperparametrów modelu Random Forest.

Hiperparametr	Zakres wartości
model__n_estimators	randint(200, 1740)
model__max_depth	randint(5, 30)
min_samples_split	randint(2, 20)
min_samples_leaf	randint(1, 10)
max_features	sqrt, log2
bootstrap	(gini, entropy)

# TUNOWALNOŚĆ ALGORYTMU

Dataset	Metoda Optymalizacji	
	Random Search	Bayes Search
diabeties (37)	0.000859	0.0021
creditg (31)	0.000419	0.00094
spambase (44)	0.00212	0.0010
yeast (40597)	0.00601	-0.0214

Tabela 6: Tabela tunowalności dla *RandomForest*.



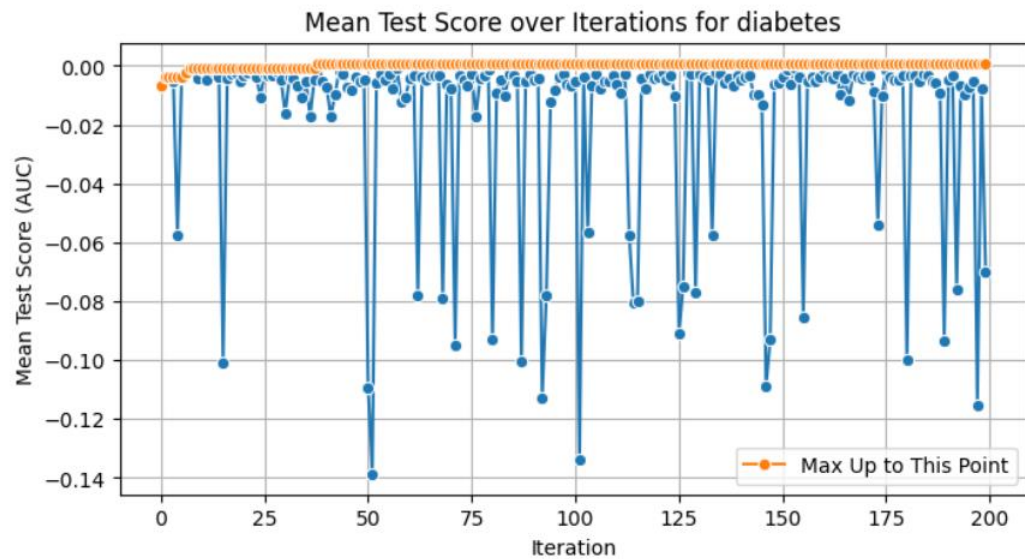
# TUNOWALNOŚĆ HIPERPARAMETRÓW

Dataset	n_estimators	max_depth	min_samples_split	min_samples_leaf	max_features	bootstrap	criterion
diabetes (37)	-0.0330	-0.0326	-0.0300	-0.0339	-0.0333	-0.0326	-0.0344
creditg (31)	0.0219	0.0222	0.0249	0.0209	0.0215	0.0223	0.0205
spambase (44)	-0.1747	-0.1743	-0.1717	-0.1756	-0.1751	-0.1743	-0.1761
yeast (40597)	-0.0027	-0.0023	0.0003	-0.0024	-0.0195	-0.0027	-0.0024

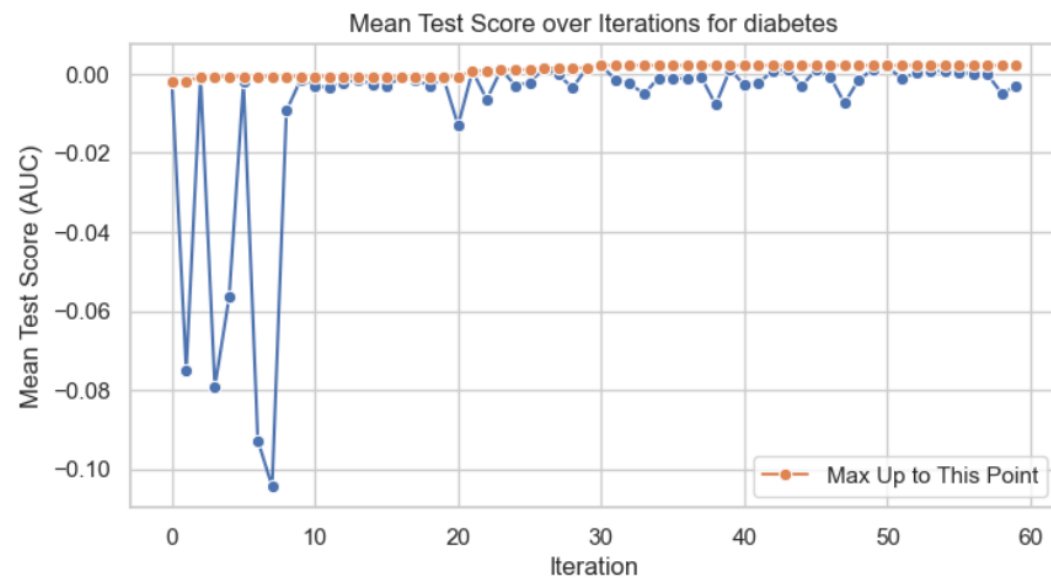
Tabela 9: Tunowalność hiperparametrów dla metody *Random Search* dla *Random Forest*.



# ZBIEŻNOŚĆ



Rysunek 9: Zbieżność dla metody *Random Search* przy użyciu `RandomForestClassifier`, dane **diabetes**.



Rysunek 10: Zbieżność dla metody *Bayes Search* przy użyciu `RandomForestClassifier`, dane **diabetes**.





KONIEC

