



15.11.2024 r.

Tunability

**Badanie tunowalności algorytmów
uczenia maszynowego**

Natalia Choszczyk | Filip Langiewicz



Agenda

03

Zbiory danych

04

Modele

05

Sampling

06

Wyniki

07

Tunowalność

08

Iteracje

09

Wnioski

10

Pytania



ZBIORY DANYCH



Credit Approval (credit)

- 690 obserwacji
- 16 kolumn
- dane kategoryczne i numeryczne
- zawiera braki danych



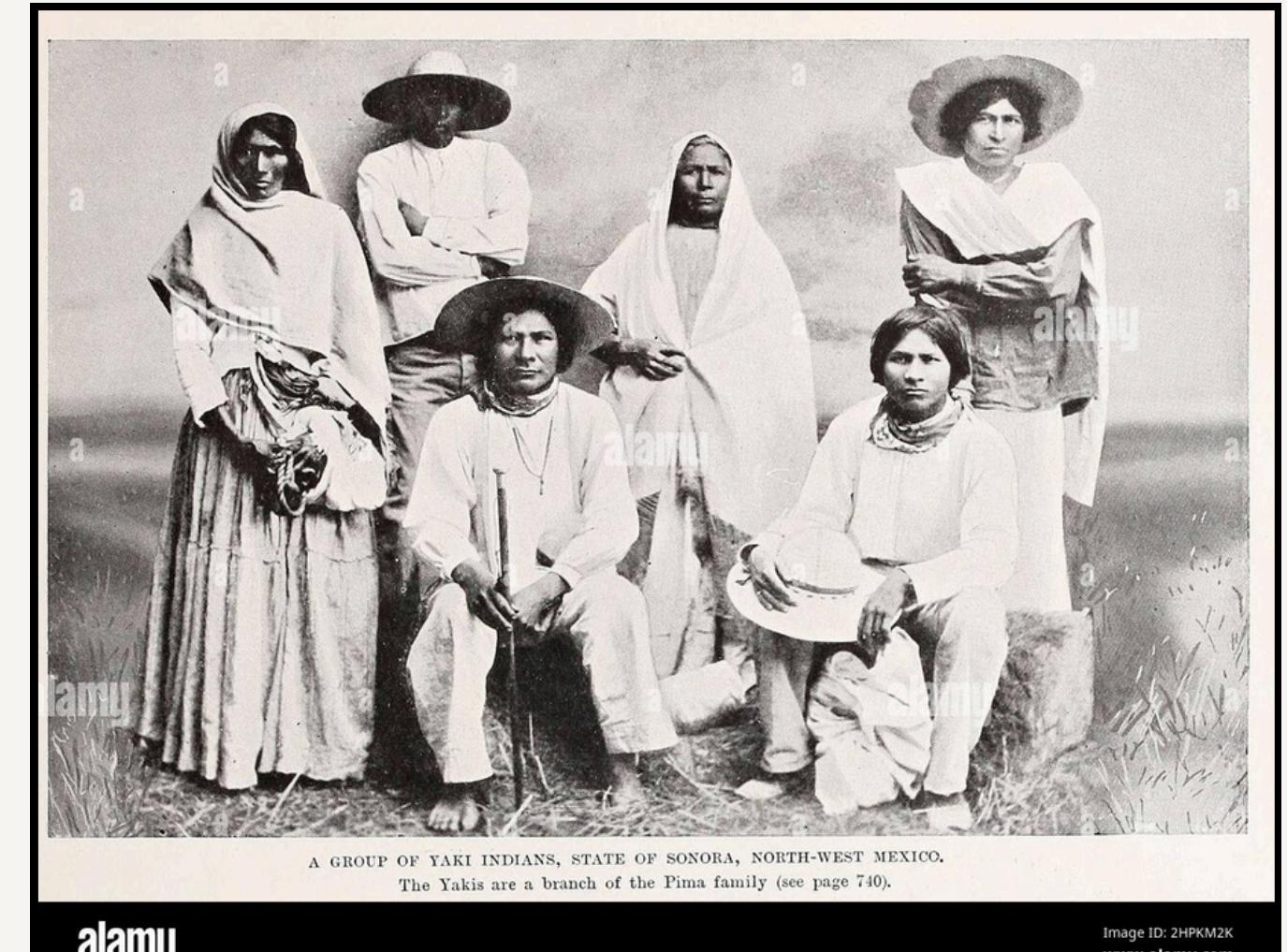
Go To College (college)

- 1000 obserwacji
- 11 kolumn
- dane kategoryczne i numeryczne
- nie zawiera braków danych
- zbiór syntetyczny



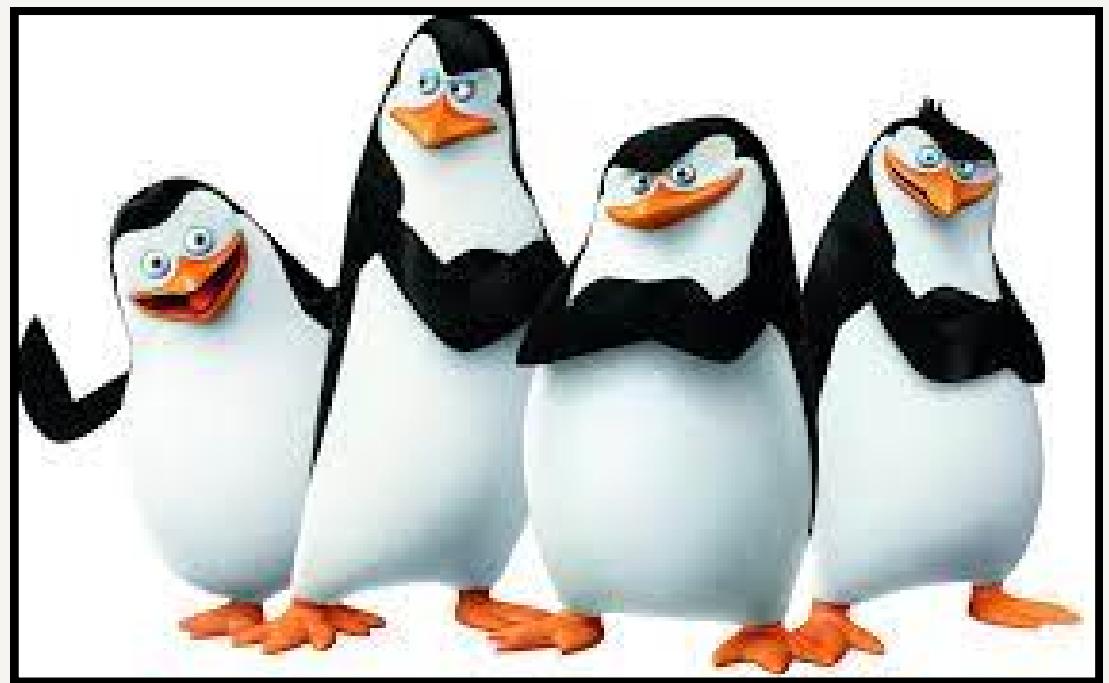
Pima Indians Diabetes Database (diabetes)

- 768 obserwacji
- 9 kolumn
- dane numeryczne
- zawiera braki danych

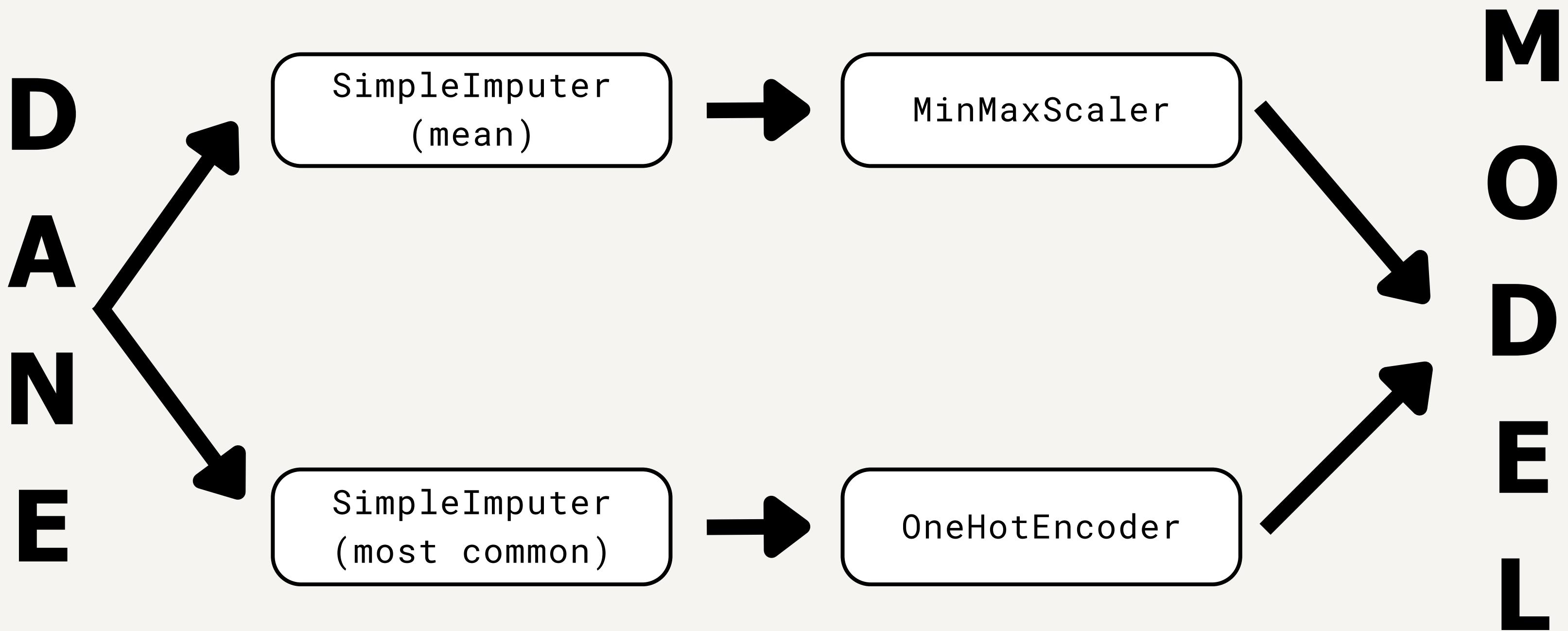


Palmer Penguins (penguins)

- **274 obserwacje**
- **7 kolumn**
- **dane numeryczne i kategoryczne**
- **nie zawiera braków danych**



Pipeline



MODELE



DecisionTree



- **criterion**: [`'gini'`, `'entropy'`, `'log_loss'`]
- **max_depth** (int): [1, ..., 29]
- **max_leaf_nodes** (int): [10, 20, ..., 90]



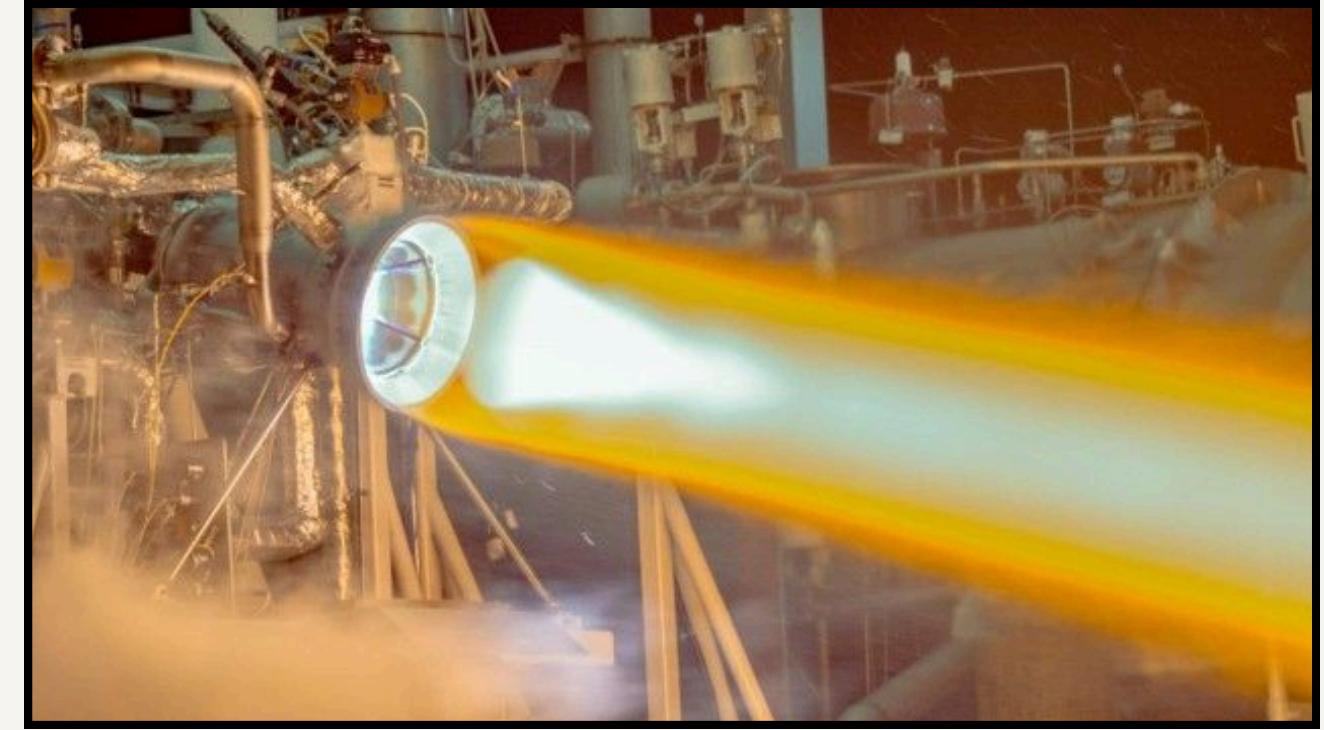
KNeighbors

- **n_neighbors** (int): [1, -, 30]
- **weights**: ['uniform', 'distance']
- **metric**: ['euclidean', 'manhattan', 'chebyshev',
'minkowski']



XGBoost

- **max_depth** (int): [3, 10]
- **learning_rate** (numeric): [0.0001, 0.3]
- **lambda** (numeric): [2^{-10} , 2^{10}]
- **booster**: ['gbtree', 'gblinear', 'dart']



RANDOM SEARCH

VS

BAYES SEARCH



RANDOM SEARCH

zbiór danych \ model	DecisionTree	KNeighbors	XGBoost
credit	0.86	0.85	0.86
college	0.86	0.85	0.90
diabetes	0.63	0.66	0.65
penguins	1.00	1.00	0.99

zbiór danych \ model	DecisionTree	KNeighbors	XGBoost
credit	0.86	0.84	0.86
college	0.86	0.85	0.90
diabetes	0.65	0.66	0.66
penguins	1.00	1.00	0.99

BAYES SEARCH



TUNABILITY



Algorytm	Zbiór danych	Tunowalność dla Random Search	Tunowalność dla Bayes Search
Decision Tree	credit	0.0254	0.0254
	college	0.0073	0.0091
	diabetes	0.0000	0.0140
	penguins	0.0000	0.0034
KNeighbors	credit	0.0064	0.0031
	college	0.0262	0.0262
	diabetes	0.0000	0.0000
	penguins	0.0069	0.0069
XGBoost	credit	0.0134	0.0134
	college	0.0000	0.0074
	diabetes	0.0000	0.0115
	penguins	0.0000	0.0101



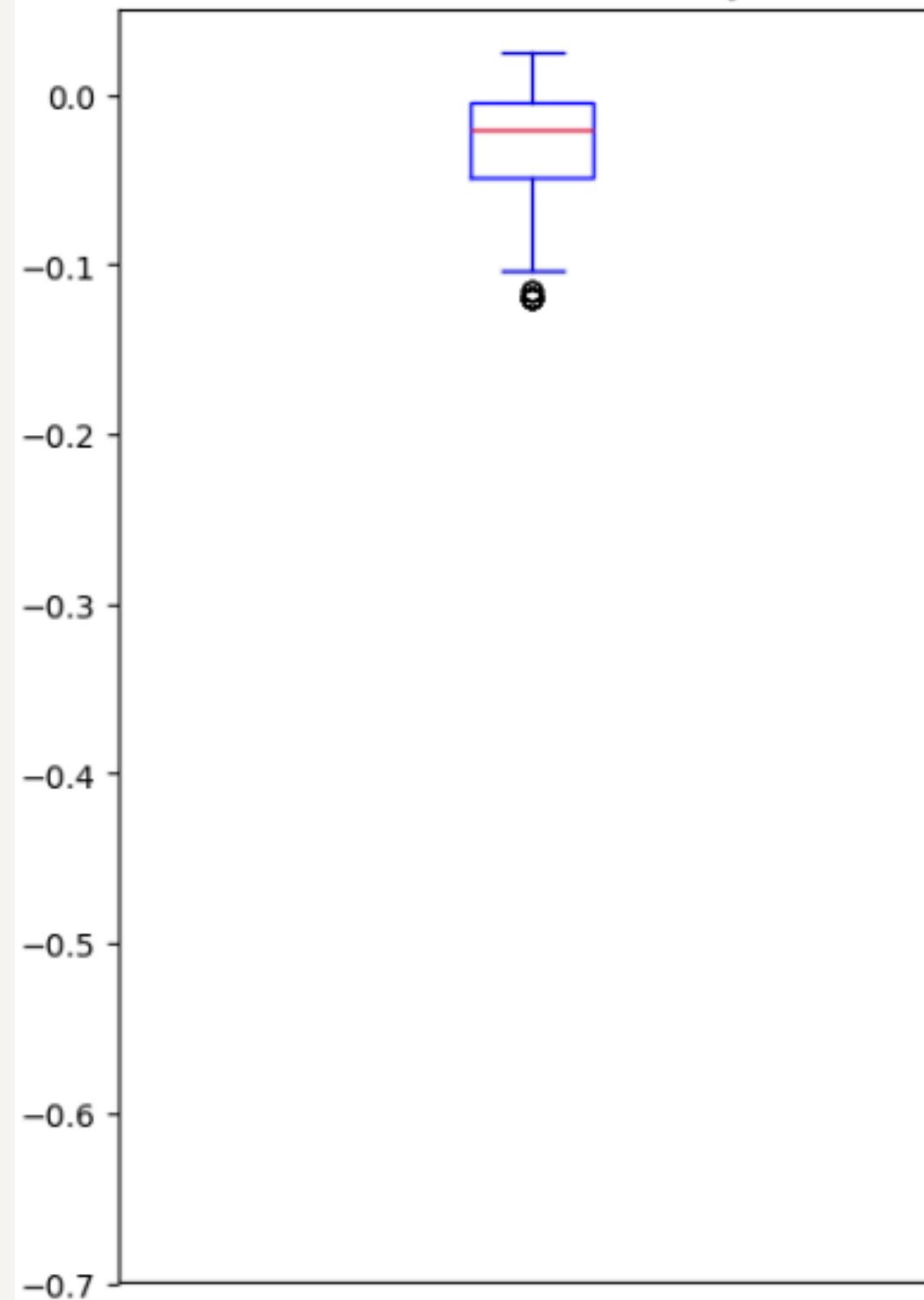
Tunowalność algorytmów

metoda samplingu \ algorytm	<u>DecisionTree</u>	KNeighbors	XGBoost
Random Search	0.0082	0.0099	0.0033
Bayes Search	0.0130	0.0091	0.0106

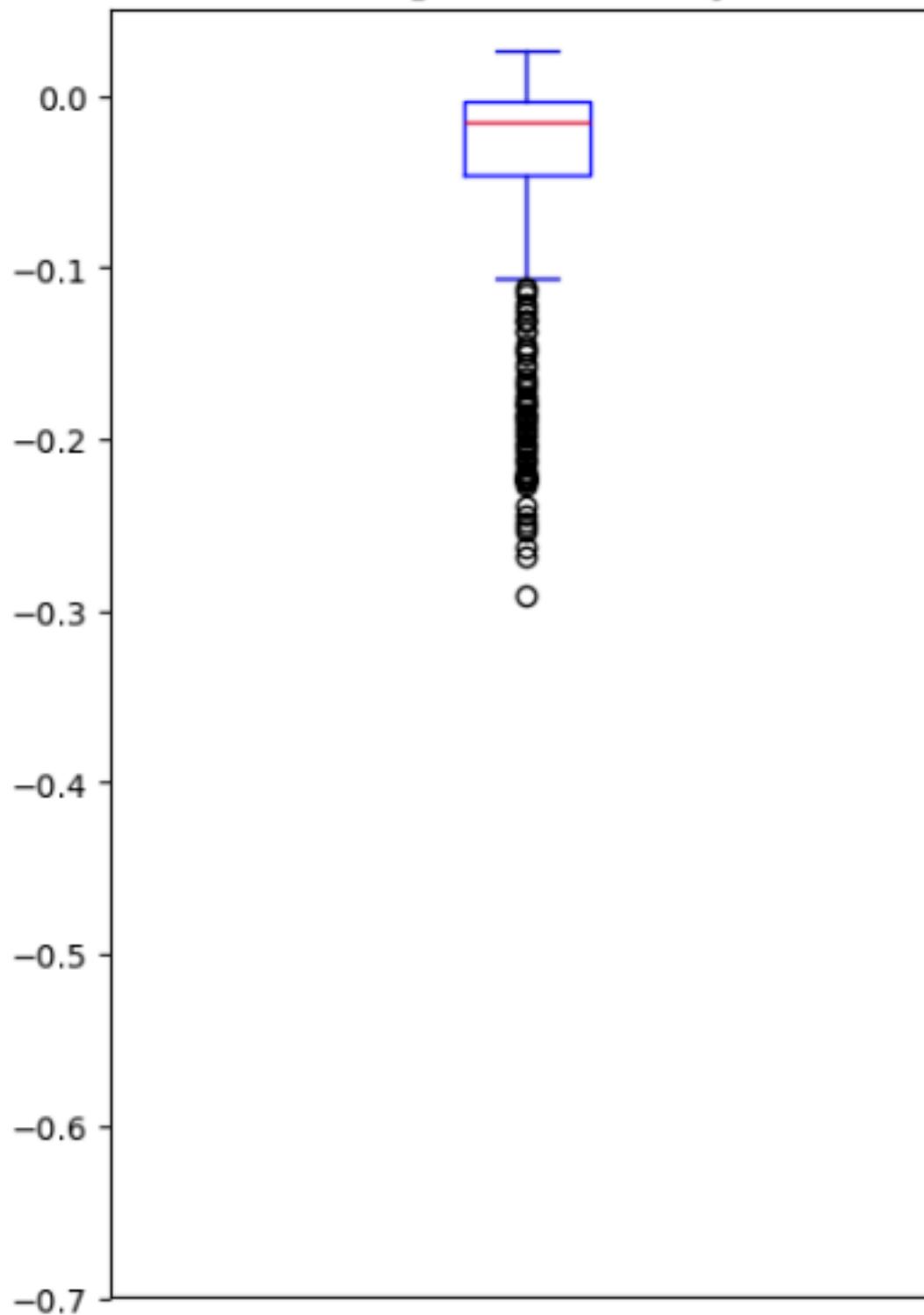


Models tunability

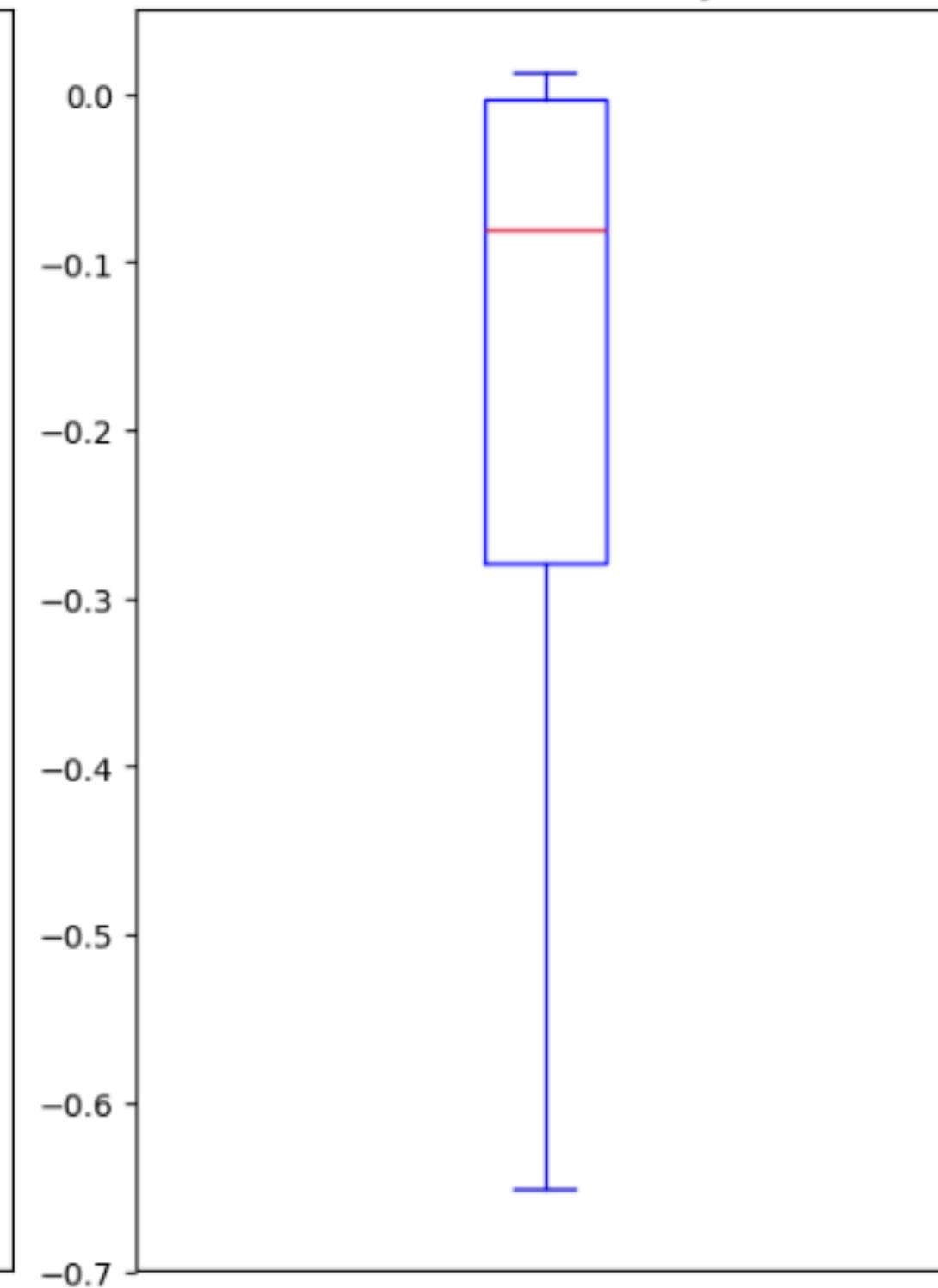
DecisionTree tunability



KNeighbors tunability



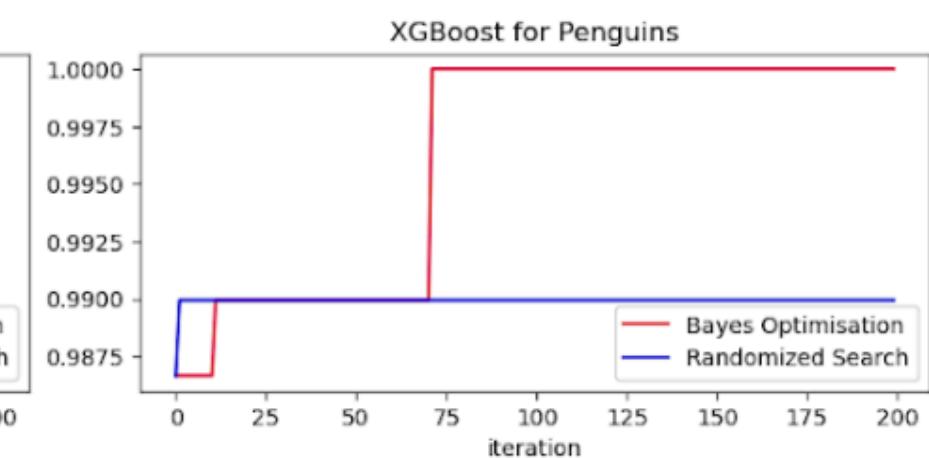
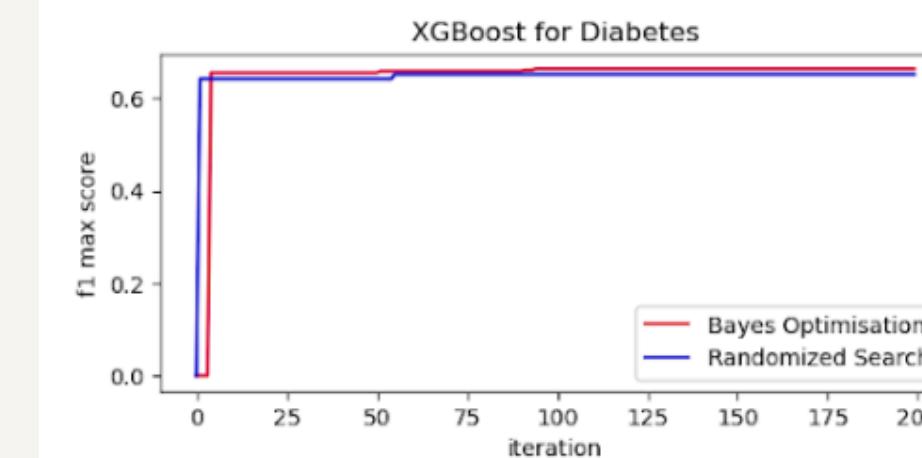
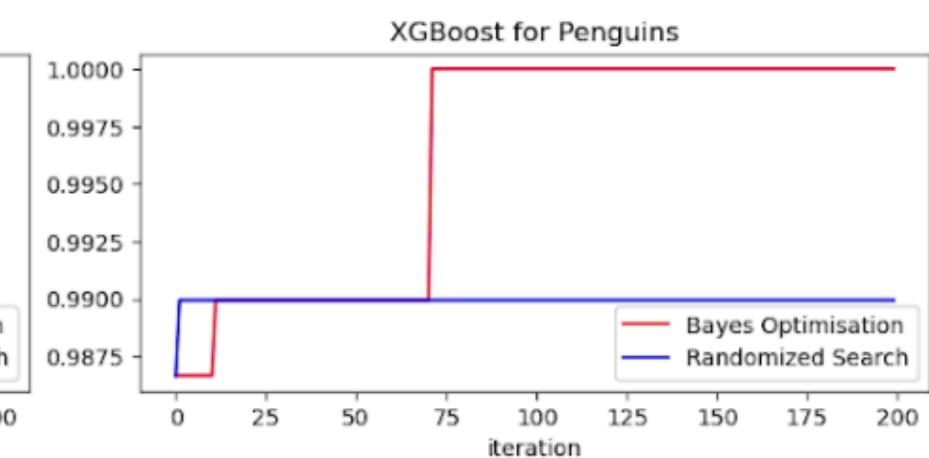
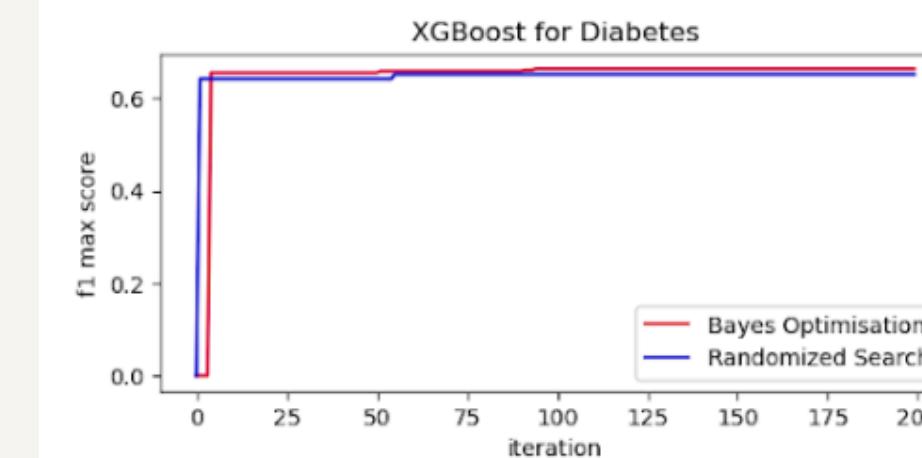
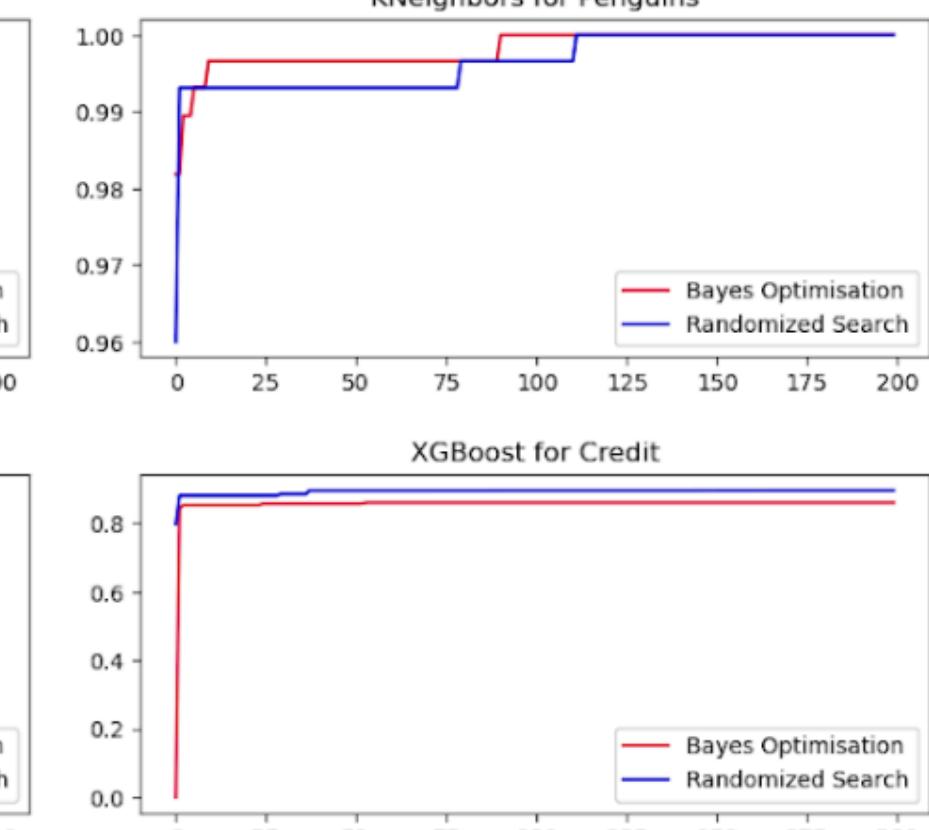
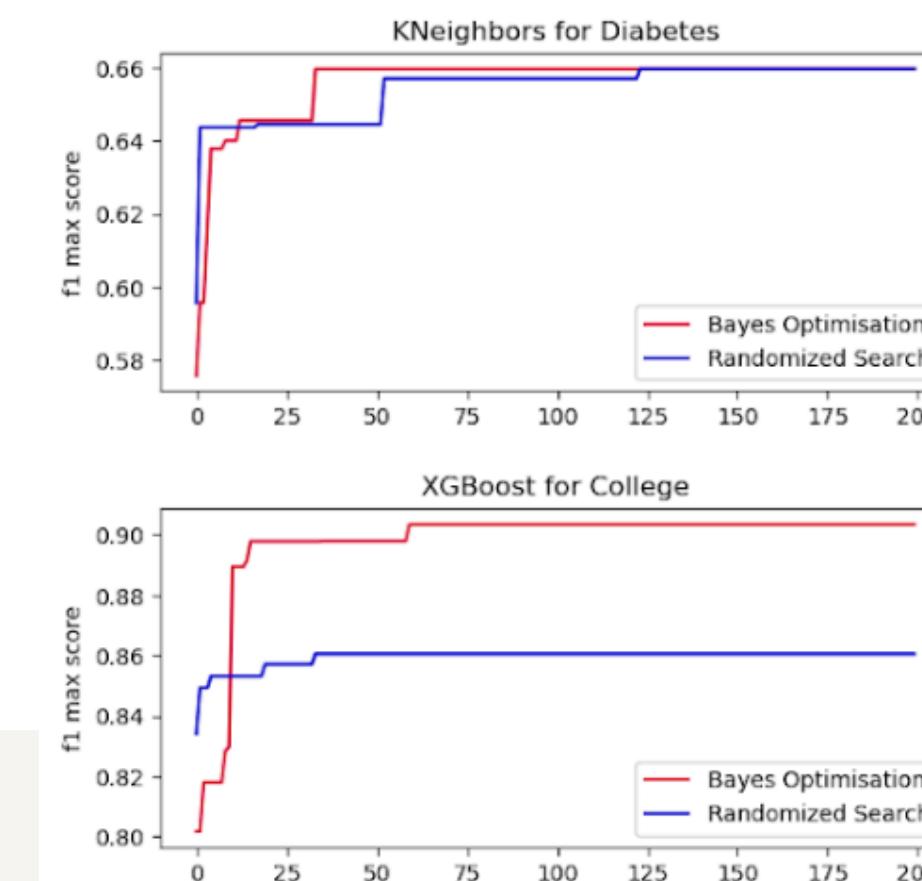
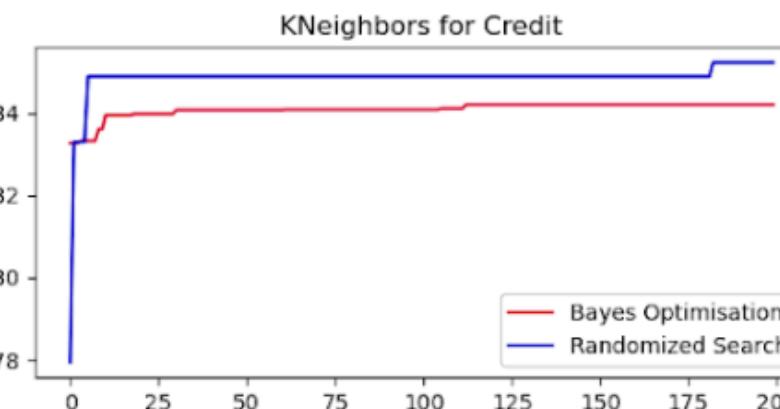
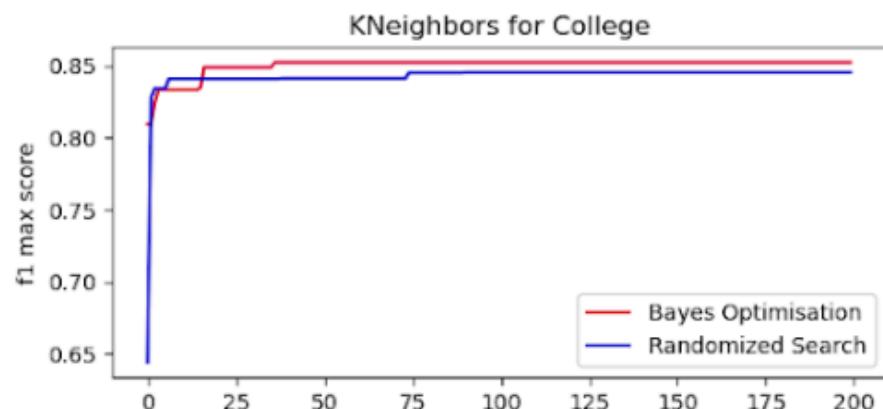
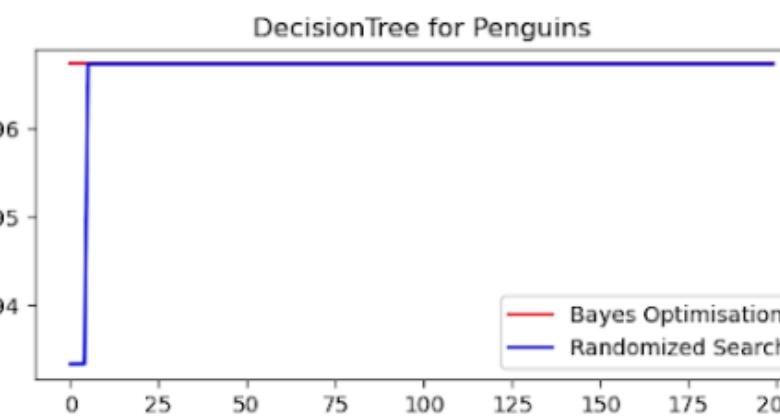
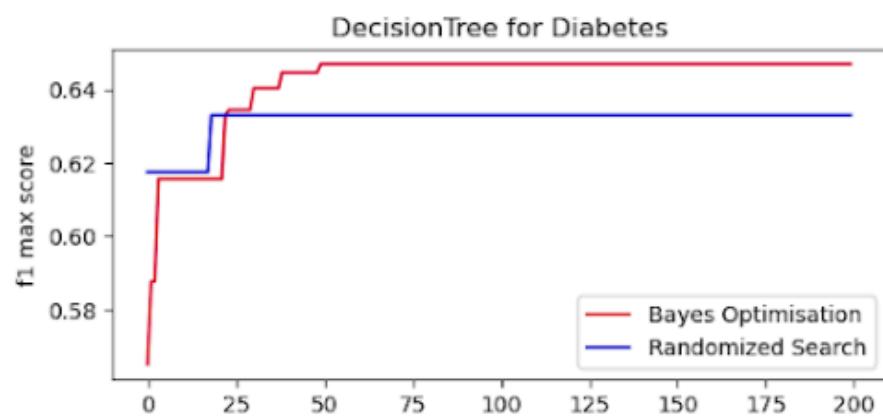
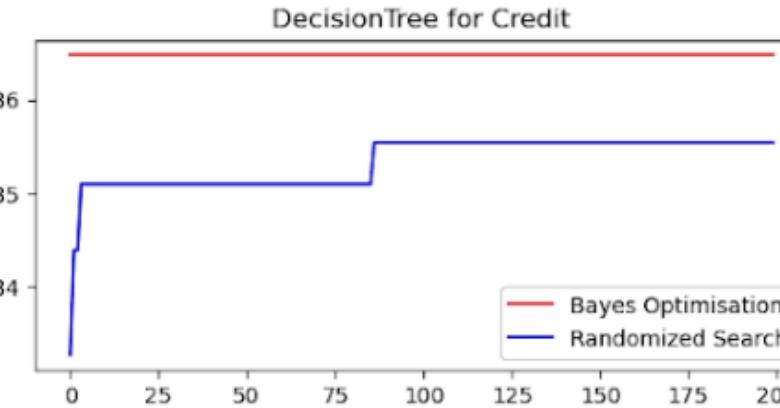
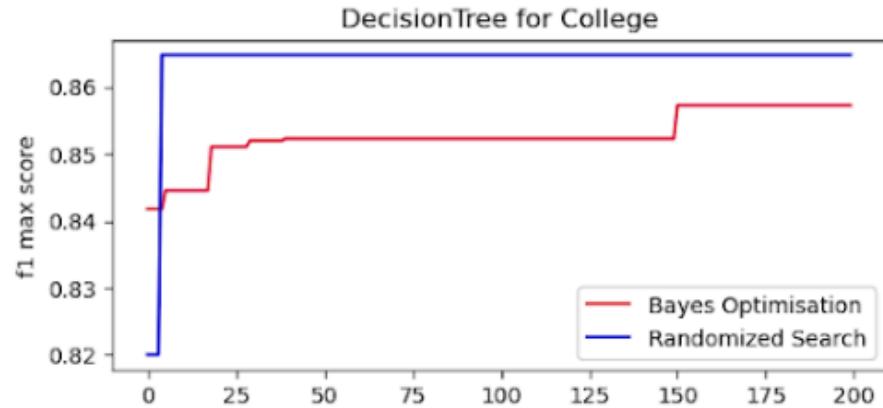
XGBoost tunability



ITERACJE



Comparision of Bayes Optimisation and Randomized Search



WINIOSKI



PYTANIA?



DZIĘKUJEMY

