



PRACA DOMOWA 1

PRZYGOTOWAŁY:

ADRIANNA WIECZOREK

JOANNA WITCZAK

SPIS TREŚCI

1. Użyte algorytmy
2. Wybrane hiperparametry
3. Wyniki
4. Wnioski

UŻYTE ALGORYTMY

- Las losowy
- Gradient boosting
- SVM

HIPERPARAMETRY DLA LASU LOSOWEGO

Dla lasu losowego rozważaliśmy następujące parametry dla metody Random Search:

- `n_estimators`, czyli liczba drzew decyzyjnych w lesie losowym,
- `min_samples_split`, czyli parametr informujący o minimalnej wymaganej liczbie obserwacji w danym węźle drzewa decyzyjnego, aby nastąpił podział węzła,
- `warm_start`, czyli parametr odpowiadający za to, czy w kolejnej iteracji używamy rozwiązania z poprzedniego dopasowania i do tego dokładamy kolejne drzewa (wartość `True`), czy dopasowujemy zupełnie nowy las (wartość `False`),
- `max_depth`, czyli maksymalna głębokość drzewa (najdłuższa ścieżka między węzłem korzenia a węzłem liścia),
- `min_samples_leaf`, czyli minimalna liczba próbek, które powinny być obecne w węźle liścia.

HIPERPARAMETRY DLA GRADIENT BOOSTING

Dla Gradient Boosting rozważaliśmy następujące parametry dla metody Random Search:

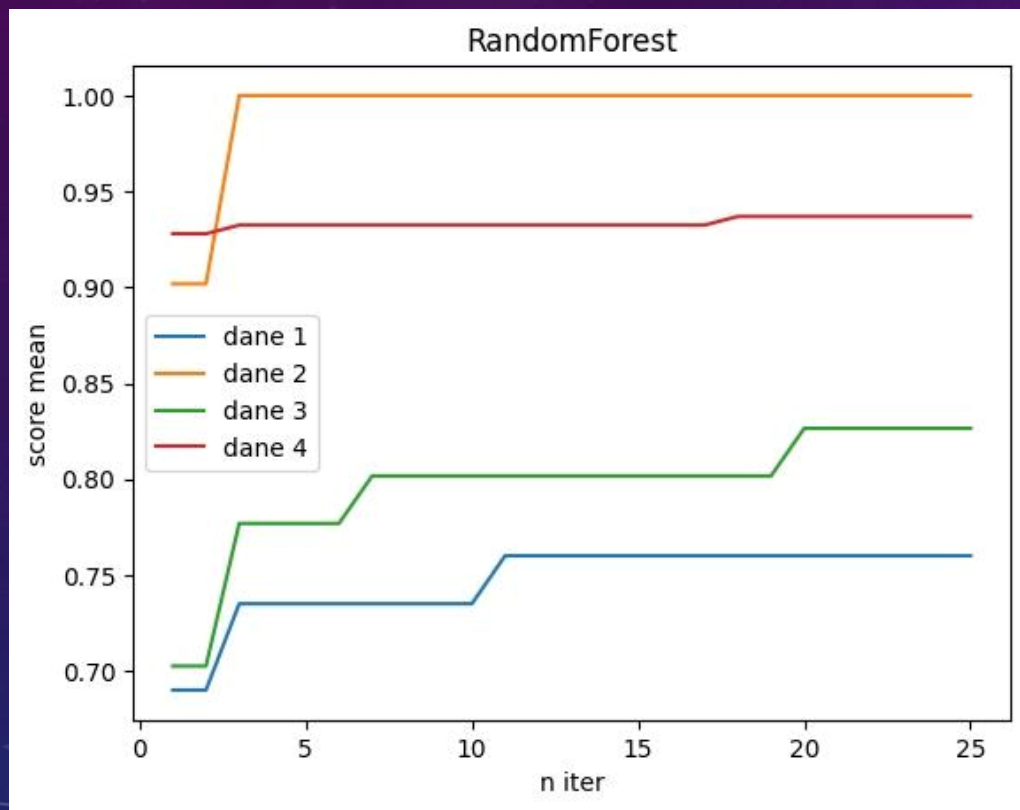
- `n_estimators`, czyli liczba drzew decyzyjnych w lesie losowym,
- `subsample`, czyli frakcja obserwacji, która ma zostać wykorzystana do dopasowania poszczególnych uczących się modeli,
- `loss`, czyli funkcja straty do optymalizacji,
- `max_depth`, czyli maksymalna głębokość drzewa (najdłuższa ścieżka między węzłem korzenia a węzłem liścia),
- `min_samples_leaf`, czyli minimalna liczba próbek, które powinny być obecne w węźle liścia.

HIPERPARAMETRY DLA SVM

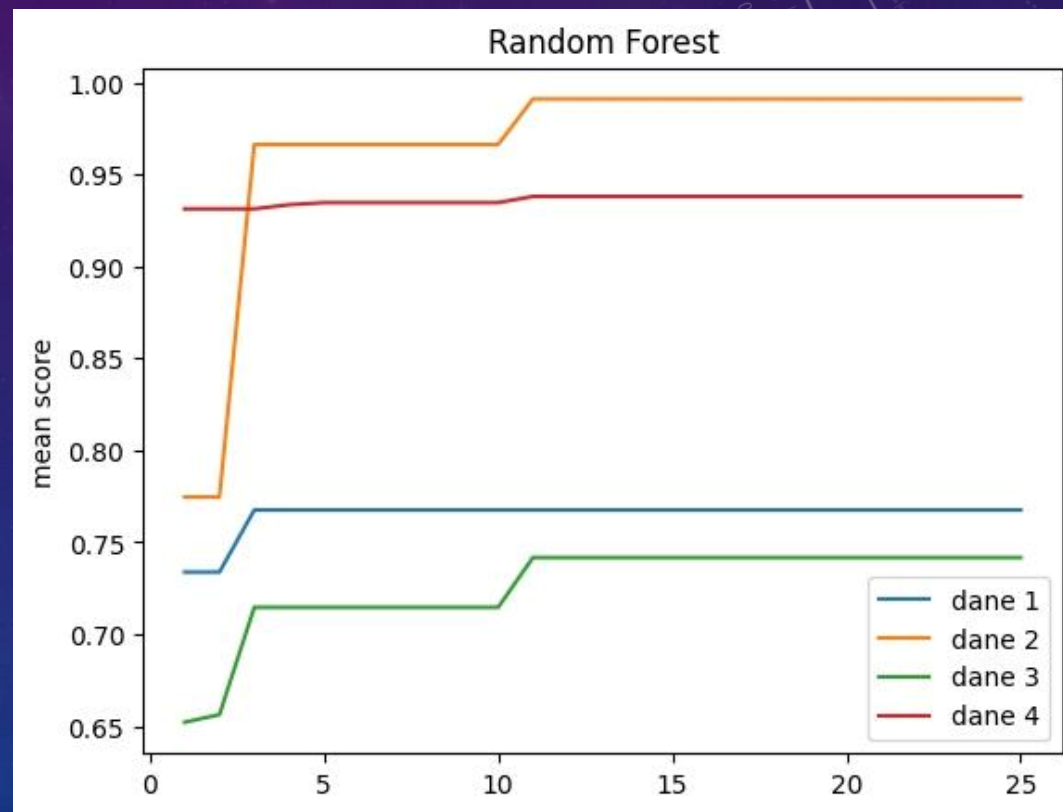
Dla SVM rozważaliśmy następujące parametry dla metody Random Search:

- C , czyli parametr regularyzacji, gdzie siła regularyzacji jest odwrotnie proporcjonalna do C ,
- kernel, czyli parametr określający typ jądra, jakie ma zostać użyte w algorytmie,
- gamma, czyli współczynnik jądra dla jąder typu „rbf”, „poly” i „sigmoid”,
- tol, czyli tolerancja dla kryterium zatrzymania.

STABILIZACJA RANDOM FOREST

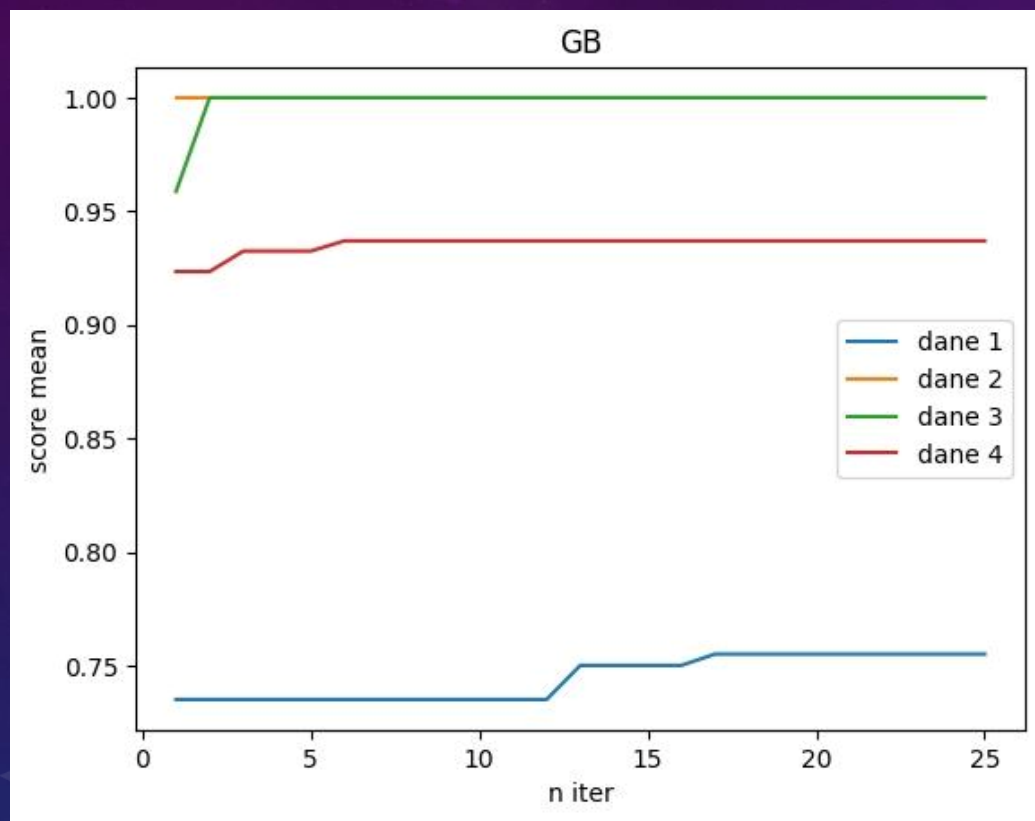


Bayes optimization

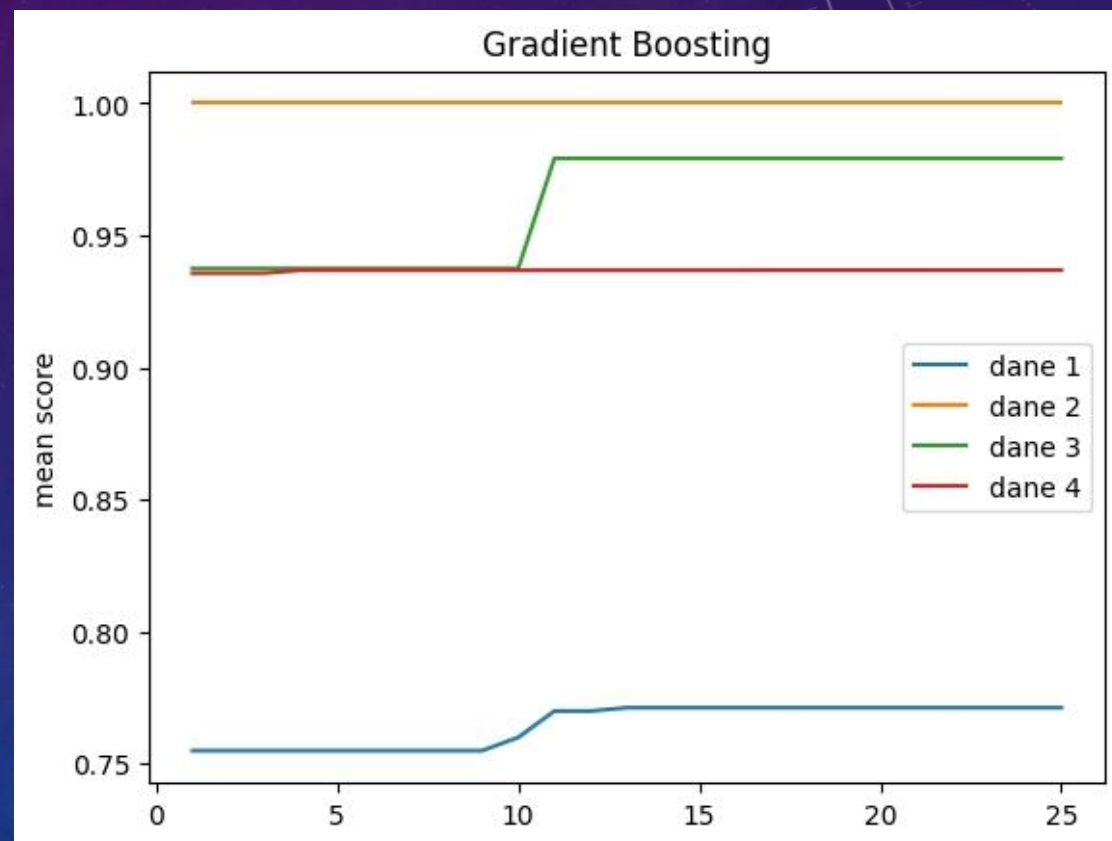


Random search

STABILIZACJA GRADIENT BOOSTING

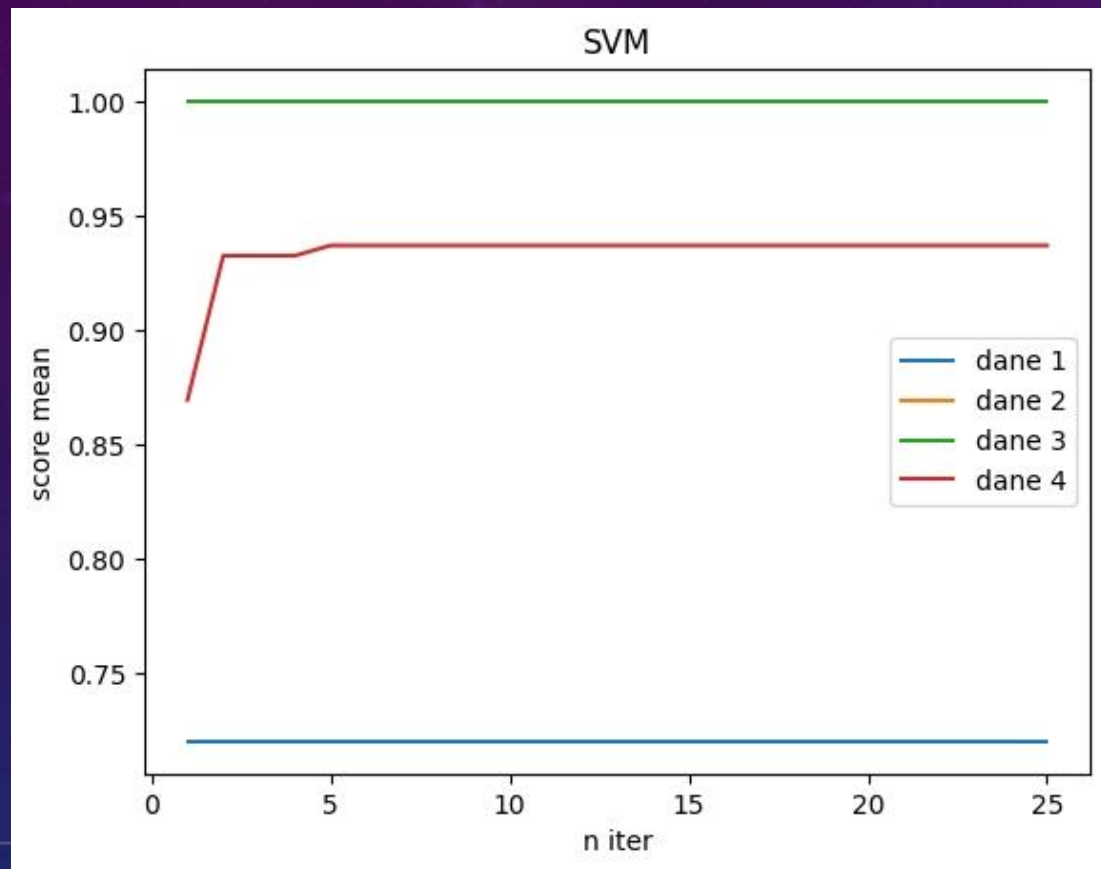


Bayes optimization

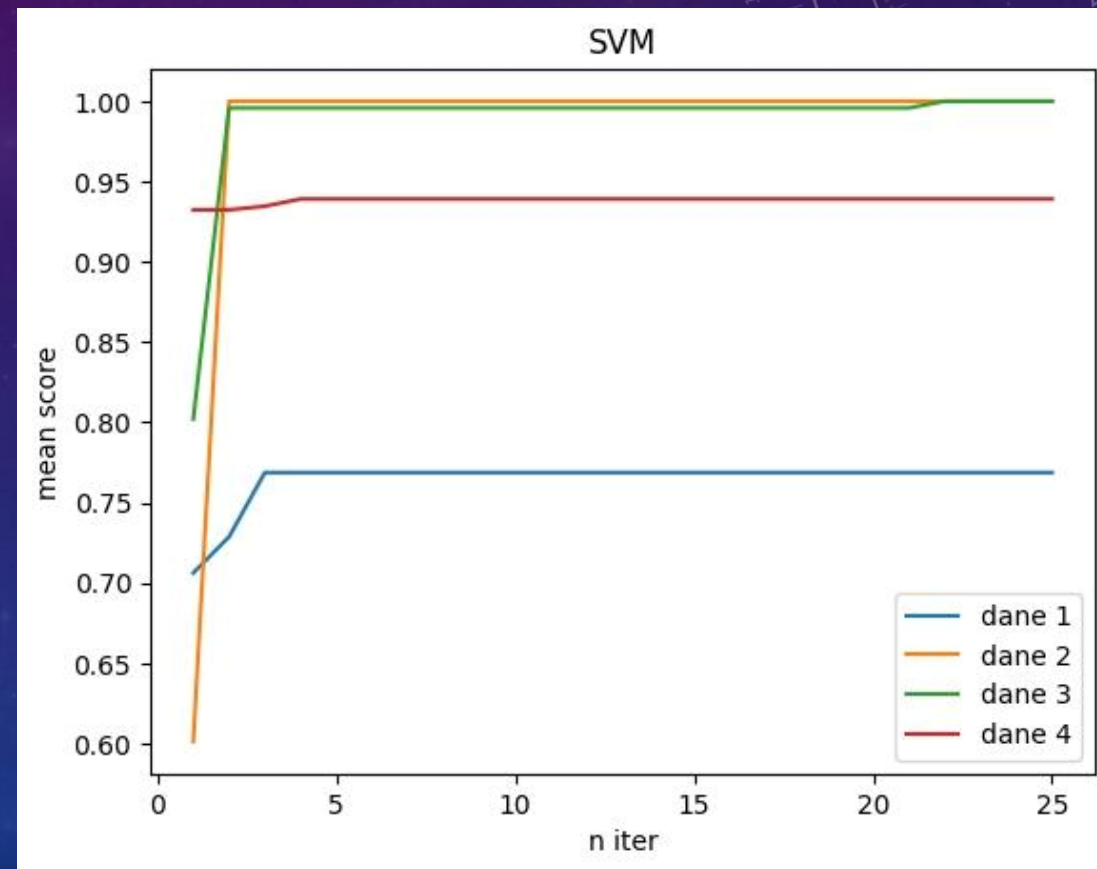


Random search

STABILIZACJA SVM



Bayes optimization

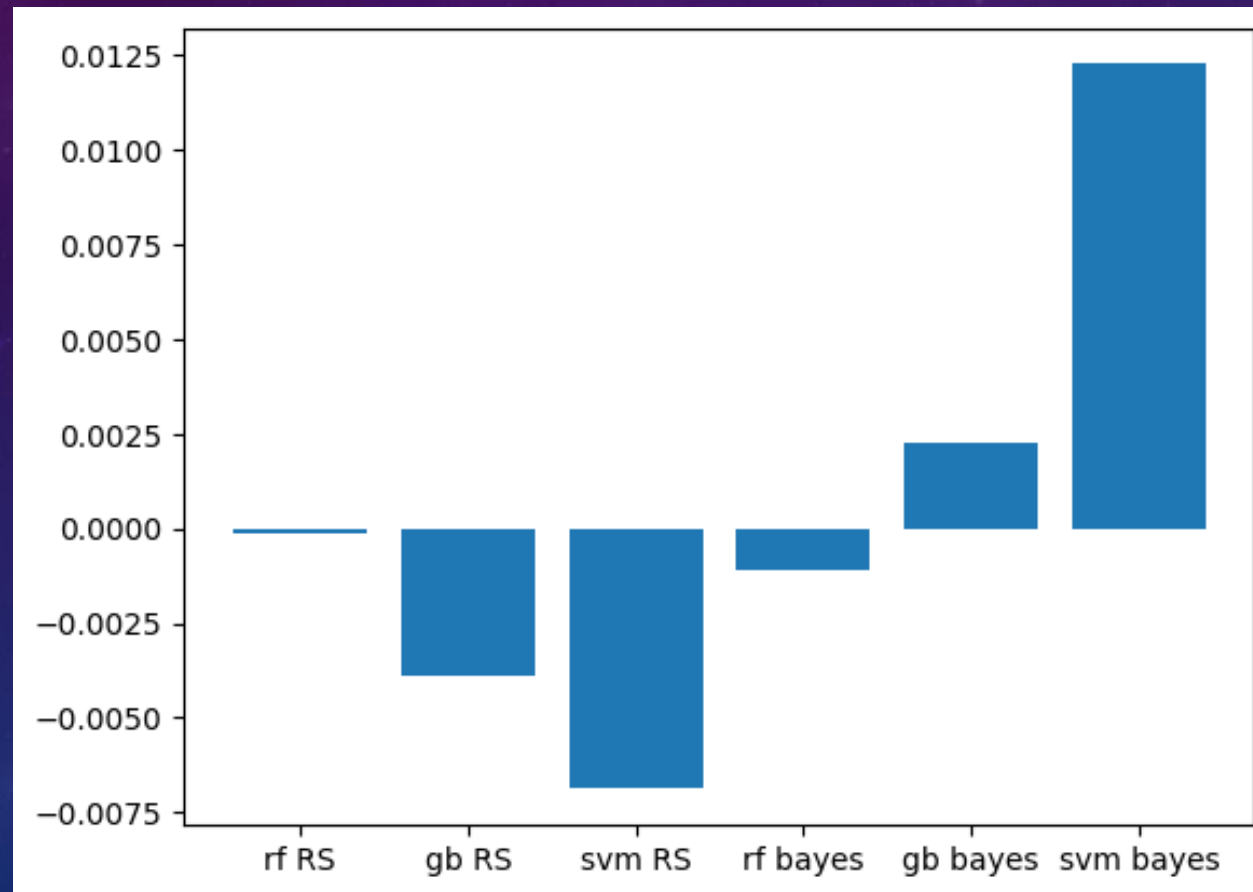


Random search

WYNIKI TESTÓW STATYSTYCZNYCH

Model	Dane	p-wartość (K-S)	p-wartość (Wilcoxon)
RandomForest	dane 1	0	0
RandomForest	dane 2	0.00004	0.0005
RandomForest	dane 3	0	0
RandomForest	dane 4	0	1.0
SVM	dane 1	0	0
SVM	dane 2	0	0.00005
SVM	dane 3	0	0.00382
SVM	dane 4	0	1
GradientBoosting	dane 1	0	0
GradientBoosting	dane 2	0.00192	0.0008
GradientBoosting	dane 3	0	0
GradientBoosting	dane 4	0	1

TUNOWALNOŚĆ ALGORYTMÓW



WNIOSKI

- Optymalizacja Bayesowska zbiega szybciej niż RandomizedSearchCV do optymalnego rozwiązania.
- Za pomocą optymalizacji Bayesowskiej otrzymujemy średnio lepsze wyniki niż za pomocą RandomizedSearchCv.t
- Badając tunowalność algorytmów zaobserwowałyśmy, że w 3 na 6 przypadków tunowalność przyjęła wartość ujemną. Sugeruje to, że dokładność otrzymana za pomocą średnio najlepszej siatki była znacząco gorsza od wyników otrzymanych za pomocą najlepszej siatki.

DZIĘKUJEMY ZA UWAGĘ :)
