Raport z analizy tunowalności hiperparametrów SVM, LightGBM , Random Forest

Julia Tomaszkiewicz, Elissa Hallak, Liwia Jankowska

Listopad 2025

1 SVM (kernel = 'rbf')

1.1 Stabilizacja optymalizacji

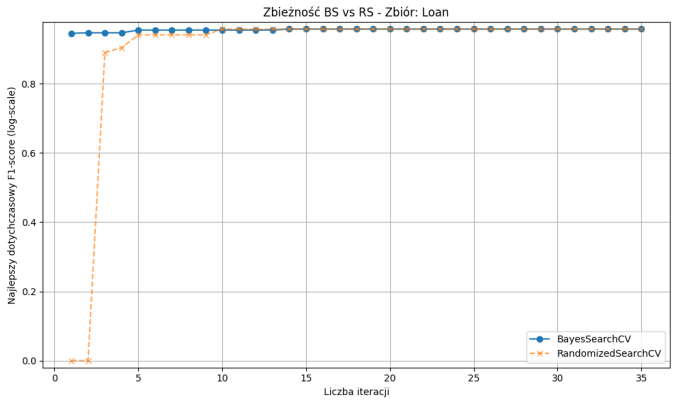


Figure 1: Analiza potrzebnych iteracji do ustabilizowania optymalizacji dla datasetu 'Loan'

Jak widać na wykresie obok, mimo że wynik rósł aż do 14. iteracji, to jest to wzrost na tyle mały, że można go zauważyć dopiero w skali logarytmicznej. Tak naprawdę już po pierwszych 5 iteracjach wynik stał się prawie zupełnie stabilny dla obu metod (sytuacja wygląda podobnie dla pozostałych datasetów [Dodatek: Fig.8]).

1.2 Przestrzeń hiperparametrów

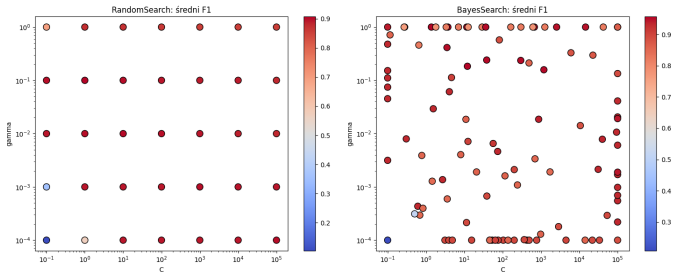


Figure 2: Porównanie Random Search i Bayes Search

Optymalizowane były dwa hiperparametry — C oraz gamma. Znaleziona literatura [1] sugerowała wartości hiperparametrów: $C = 2^{-5}, 2^{-3}, \ldots, 2^{15}$ oraz $\gamma = 2^{-15}, 2^{-13}, \ldots, 2^{3}$. Wybrane przedziały $C \in (1^{-1}, 1^{5})$ oraz $\gamma \in (1^{-4}, 1)$ zawierają się w nich, a GridSearch oraz RandomSearch wykonano na wartościach rosnących wykładniczo (na mniejszej liczbie iteracji ze względu na ograniczenia sprzętowo–czasowe). Na wykresie obok widać średnie F1 osiągane dla danej pary hiperparametrów we wszystkich datasetach (warto zauważyć, że dla Bayes Search wiele par hiperparametrów nie powtórzyło się w więcej niż 1 datasecie, ponieważ nie ma on ustalonej siatki punktów)(dodatkowe wykresy pokazują też wariancje [Dodatek: Fig.5, Fig.6]).

1.3 Tunowalność algorytmu SVM i Bias Sampling

W tabelce widzimy podsumowanie różnic najlepszego F1 osiągniętego przez daną metodę na danym datasecie, a F1 osiągniętego przez nową, średnio najlepszą (dla GridSearch) defaultową parę hiperparametrów - C: 1.0, gamma: 0.1. Co ciekawe podstawowy default [2], który zostałby wyznaczony przez SVM wynosiłby praktycznie tyle samo dla datasetu 'Personality' (C: 1.0, gamma: 0.01286).

W obu metodach datasety 'Personality' i 'Heart' pokazują kompletny brak tunowalności, 'Machine' ma niską tunowalność, a 'Loan' wypada w obu metodach najlepiej i ma najwyższą tunowalność z poprawą o 2.4 punkty procentowe. Na tej podstawie możemy stwierdzić, że technika losowania punktów nie wpływa na różnice we wnioskach dotyczących tunowalności SVM na użytych datasetach. Bias Sampling występuje w BayesSearch, w którym próbkowanie jest zależne od poprzednich wyników, co nie dzieje się w GridSearch i RandomSearch.

dataset	best_f1	default_f1	tunability	method
Personality	0.931912	0.931912	0.000000	GridSearch
Heart	0.865508	0.865508	0.000000	GridSearch
Loan	0.957788	0.933607	0.024180	GridSearch
Machine	0.904952	0.899395	0.005557	GridSearch
Personality	0.931912	0.931912	0.000000	BayesSearch
Heart	0.862270	0.865508	-0.003238	BayesSearch
Loan	0.957841	0.933607	0.024233	BayesSearch
Machine	0.906771	0.899395	0.007375	BayesSearch

Figure 3: Tunowalność SVM dla różnych datasetów

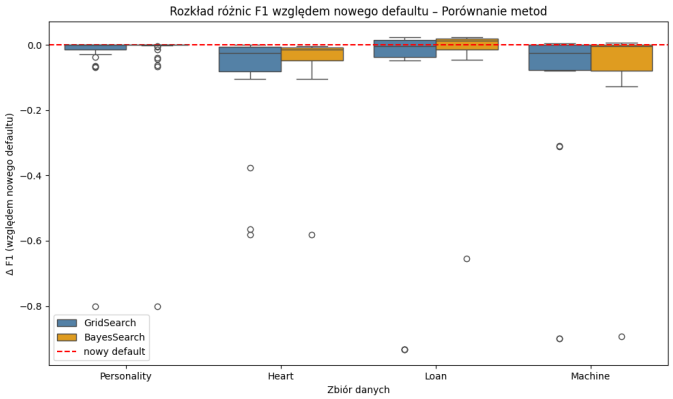


Figure 4: Boxploty pokazujące różnicę między F1 dla każdej pary hiperparametrów, a F1 naszego defaultu (wersja z RandomSearch w [Dodatek: Fig.7]

2 LightGBM

Hiperparametr	**Random Search**	**BayesSearchCV**
`num_leaves`	`randint(20, 150)`	`Integer(20, 150)`
`max_depth`	`randint(8, 15)`	`Integer(1, 15)`
`learning_rate`	`loguniform(0.001, 0.3)`	`Real(0.001, 0.3, log-uniform)`
`n_estimators`	`randint(50, 5000)`	`Integer(50, 5000)`
`min_child_samples`	`randint(1, 100)`	`Integer(1, 100)`
`subsample`	`uniform(0.6, 0.4)`	`Real(0.1, 1.0)`
`colsample_bytree`	`uniform(0.6, 0.4)`	`Real(0.1, 1.0)`

W załączonej tabeli przedstawione są zakresy testowanych hiperparametrów zarówno dla techniki RandomSearch jak i BayesSearch. Przedziały te zostały dobrane w oparciu o rekomendacje dostępne w literaturze, tak aby uwzględnić zarówno wartości typowe, jak i potencjalnie efektywne w praktyce.

Dodatkowo przetestowano także rozszerzony zakres dla parametru max_depth. Okazało się jednak, że zwiększenie maksymalnej głębokości drzewa nie prowadziło do poprawy jakości modelu, wyniki pozostawały praktycznie identyczne. Jednocześnie większe wartości max_depth znacząco zwiększały koszt obliczeniowy. Z tego powodu w końcowej analizie użyty został pierwotny, bardziej oszczędnym zakres.

2.1 Analiza wyników

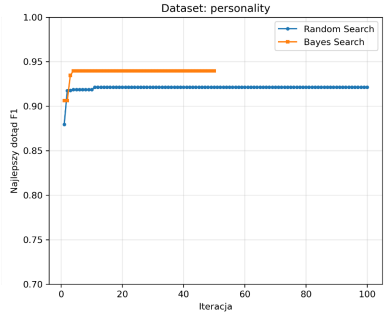


Figure 5: Analiza potrzebnych iteracji do ustabilizowania optymalizacji

Z wykresu wynika, że zarówno metoda BayesSearch, jak i Random Search wyraźnie się stabilizują. Pierwsze oznaki ustabilizowania wyników pojawiają się już po około 5 iteracjach, jednak bezpiecznym i uniwersalnym wyborem (niezależnie od zbioru danych) jest przyjęcie około 20 iteracji dla obu technik.

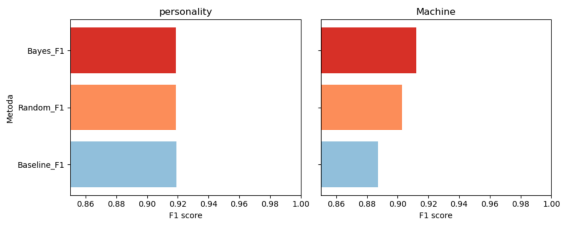
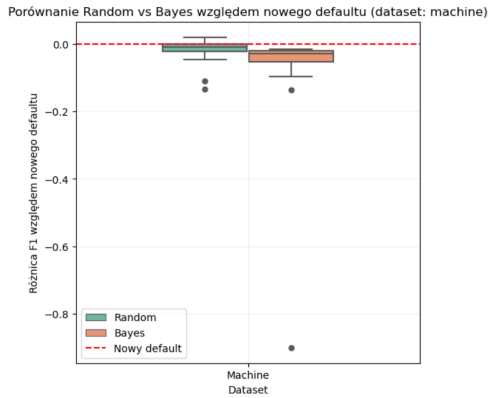


Figure 6: Najlepsze F1 osiągnięte przez GridSearch oraz RandomSearch

Analiza wykresów przedstawiających najlepsze uzyskane wartości F1 dla zbiorów danych Machine i Personality pokazuje, że różnice w jakości modeli są bardzo subtelne. Oznacza to, że przestrzeń potencjalnych ulepszeń względem konfiguracji oryginalnej (automatycznie używanej w Pythonie) jest niewielka, a oba podejścia prowadzą jedynie do nieznacznego wzrostu wartości miary F1.



Wykres przedstawia różnice wartości F1 dla wszystkich testowanych zestawów hiperparametrów w odniesieniu do nowo wyznaczonego „default". Wyraźnie widać, że znaczna część konfiguracji daje wyniki bardzo zbliżone do nowego default, co wskazuje, że model jest mało wrażliwy na dobór hiperparametrów. Większość zestawów zapewnia porównywalną wydajność. Dodatkowo można zauważyć, że zarówno BayesSearch, jak i Random Search osiągają podobne wyniki, a wybór techniki próbkowania nie wpływa znacząco na efektywność modelu. Nie występuje bias sampling.

Table 1: Porównanie wyników F1 dla Random Search i BayesSearch

dataset	best F1 Random	best F1 Bayes	default F1	tunability Bayes	tunability Random
Personality	0.918794	0.918794	0.918981	-0.00019	-0.00019
Heart	0.919614	0.915584	0.894040	0.021544	0.025574
Loan	0.975238	0.973077	0.971319	0.001758	0.003919
Machine	0.902954	0.912000	0.887029	0.024971	0.015925

Z powyższej tabeli wynika, że wartości F1 dla Random Search i BayesSearchCV są bardzo zbliżone do wartości domyślnych. Tunability w większości przypadków jest niewielka, co potwierdza wysoką efektywność domyślnych ustawień parametrów i ograniczoną możliwość istotnej poprawy wyników poprzez strojenie.

3 Random Forest

3.1 Przestrzeń hiperparametrów i Bias Sampling

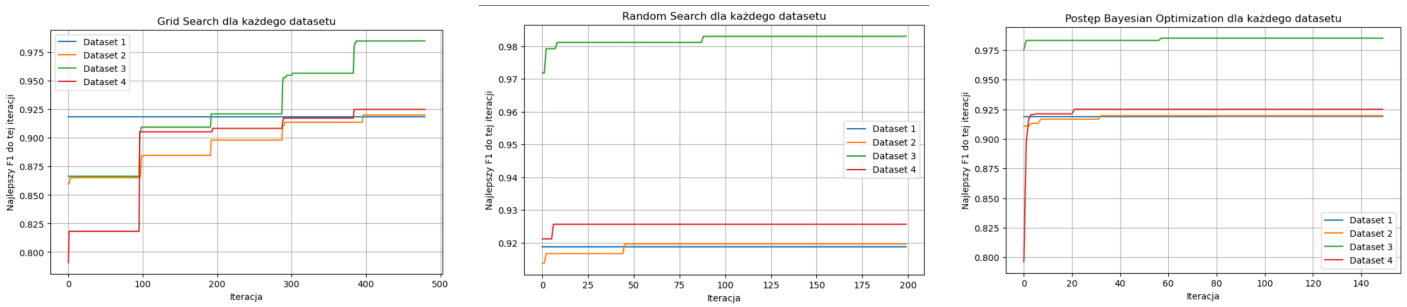
Optymalizacji poddałam 5 hiperparametrów: n estimators, max depth, min samples split, min samples leaf oraz max features. W literaturze nie są podane konkretne zakresy wartości hiperparametrów, które warto przetestować, a raczej ogólna filozofia, jaką należy się kierować, np. to, że liczba większa liczba drzew zazwyczaj daje lepsze wyniki (jednak żadne konkretne liczby nie są podawane) [3]. Poniżej przedstawiłam zakresy, jakie wybrałam dla Random Search i Grid Search

Hiperparametr	**Grid Search**	**Random Search**
`n_estimators`	[60, 100, 140, 180, 220]	`randint(60, 220)`
`max_depth`	[1, 3, 5, 7, `None`]	`random.choice([1, 3, 5, 7])`
`min_samples_split`	[2, 3, 4, 5]	`randint(2, 5)`
`min_samples_leaf`	[1, 2, 3]	`randint(1, 3)`
`max_features`	[`'sqrt'`, `'log2'`]	random.choice([`'sqrt'`, `'log2'`])

Odpowiedź na pytanie czy występuje Bias Sampling jest taka sama jak w przypadku SVM (punkt 1.3). Wykres dla zakresu przeszukanych parametrów w Random Searchu znajduje się w pliku z dodatkowymi wykresami.

3.2 Stabliność metod

Z poniższych wykresów wynika, że dla optymalizacji Bayesowskiej stabilne wyniki uzyskujemy już po 60 iteracji dla wszystkich dataset'ów, a dla Random Searcha po ok. 90 iteracji. Użyty Grid Search przechodzi przez punkty w siatce po kolei, więc ciężko jest ocenić jego "stabilność" (jako, że nie jest oparty na losowaniu punktów), jednak widać, że jedynie dla 1 zestawu danych algorytm się "stabilizuje".



(a) Grid Search (b) Random Search (c) Bayes Search

Figure 7: Porównanie stabilności metod strojenia hiperparametrów

3.3 Tunowalość algorytmu Random Forest

Table 2: Porównanie wyników F1-score z uwzględnieniem procentowej poprawy

Dataset	Grid dom.	Grid naj.	Róż. Grid [%]	Random dom.	Random naj.	Róż. Random [%]	Bayes	Róż. Bayes [%]	Base.
Dataset 1	0.91647	0.91879	0.25%	0.91879	0.91879	0.00%	0.91898	0.02%	0.89906
Dataset 2	0.91961	0.92013	0.06%	0.91667	0.91961	0.32%	0.91961	0.32%	0.91318
Dataset 3	0.98311	0.98502	0.19%	0.98311	0.98311	0.00%	0.98502	0.19%	0.98311
Dataset 4	0.91213	0.92500	1.41%	0.92050	0.92562	0.56%	0.92500	0.49%	0.92500

W powyższej tabeli przedstawiono porównanie F1 score osiąganych dla najlepszych i nowych "domyślnych" konfiguracji z poszczególnych algorytmów (wyniki z Bayesa zostały porównane z domyślym Random Search), jak i wyniki osiągane dla "oryginalnej"-domyślnej konfiguracji (kolumna "Base"). Jak widać, tunowanie hiperparametrów ma wyraźny sens tylko w przypadku datasetu 4, gdzie Grid Search przyniósł aż 1.41% poprawy (zatem ten dataset jest najbardziej tunowalny). Dla pozostałych trzech datasetów zyski z tuningu były marginalne - mieszczące się w przedziale 0.06% do 0.32%. To sugeruje, że w większości przypadków nowe domyślne parametry Random Forest są już blisko optimum i czas poświęcony na ich strojenie może nie być wart tak niewielkiej poprawy

References

[1] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin *A Practical Guide to Support Vector Classification.*
`https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf`

[2] *scikit-learn SVM github*
`https://github.com/scikit-learn/scikit-learn/blob/1eb422d6c5/sklearn/svm/_classes.py#L614`

[3] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.