

Automatyczne uczenie maszynowe – projekt 1

Magdalena Bartczak, Aleksandra Mach

29.11.2023

- 1 **Zbiór 1: *cylinder-bands*** – zbiór rozmiaru 540×38 , zawiera 18 zmiennych numerycznych i 19 zmiennych kategorycznych; zawiera braki danych w 263 obserwacjach; zmienna odpowiedzi *band_type* przyjmuje dwie wartości: 'band' oraz 'noband'.
- 2 **Zbiór 2: *adult*** – zbiór rozmiaru 48842×15 , zawiera 6 zmiennych numerycznych i 8 zmiennych kategorycznych; zmienna odpowiedzi *class* przyjmuje dwie wartości: ' $\leq 50K$ ' oraz ' $> 50K$ '.
- 3 **Zbiór 3: *labor*** – zbiór rozmiaru 57×17 ; zawiera 8 zmiennych numerycznych i 8 zmiennych kategorycznych; zawiera braki danych w 56 obserwacjach; zmienna odpowiedzi *class* przyjmuje dwie wartości: 'good' oraz 'bad'.
- 4 **Zbiór 4: *sick*** – zbiór rozmiaru 3772×30 ; zawiera 6 zmiennych numerycznych i 23 zmienne kategoryczne; zawiera braki danych w 3772 obserwacjach; zmienna odpowiedzi *Class* przyjmuje dwie wartości: 'sick' oraz 'negative'.

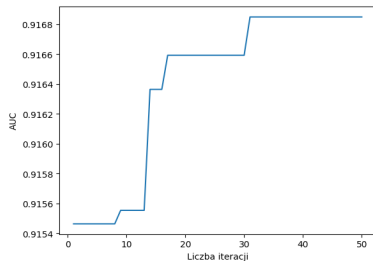
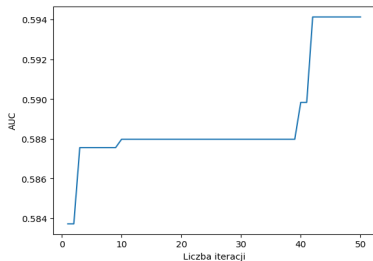
Wybrane metody samplingu:

- Random Search
- Bayesian Optimization (pakiet SMAC)

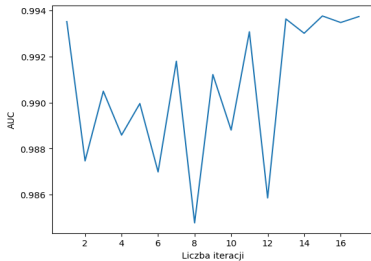
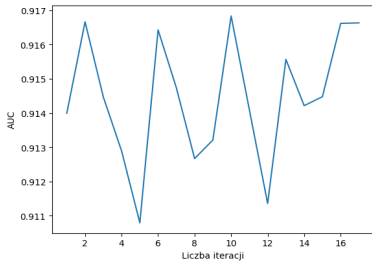
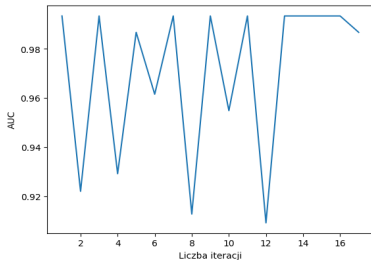
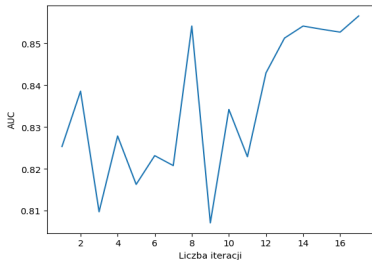
Wybrane metody uczenia maszynowego:

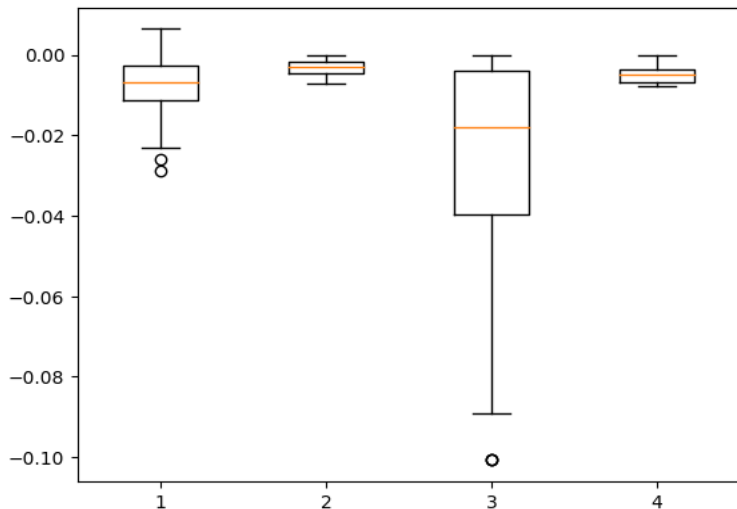
- Random Forest
- XGBoost
- SVM

RandomSearch dla lasu losowego

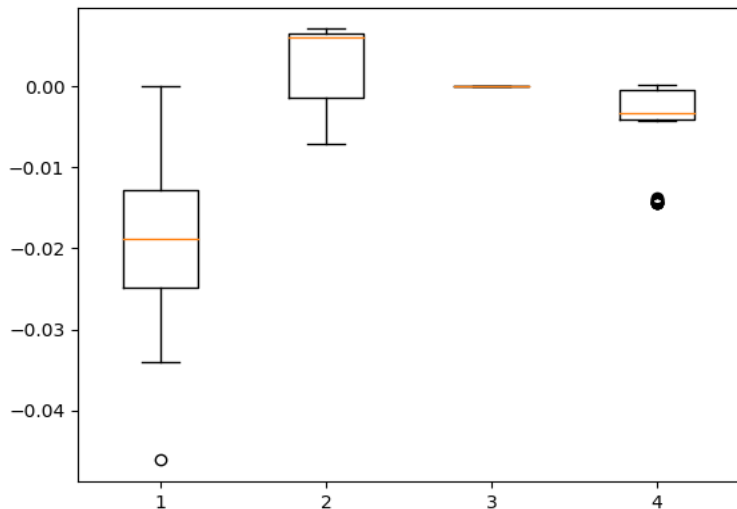


Optimalizacja bayesowska dla lasu losowego





Tuning XGBoosta



Dziękujemy za uwagę!