Analysis of Halving Random Search Cross Validation in Optimizing Machine Learning Models

Author : Alief Cahyo Utomo (aliepcahyo@gmail.com)

Affiliation : Universitas Nusantara PGRI Kediri, Jl. Ahmad Dahlan No.76, Mojoroto, Kediri, East Java, Indonesia

Background & Research Problem

Background

- Model performance heavily depends on optimal hyperparameter configuration.
- Traditional tuning methods like Grid Search CV are computationally expensive on large datasets as they evaluate all combinations.

Previous Studies

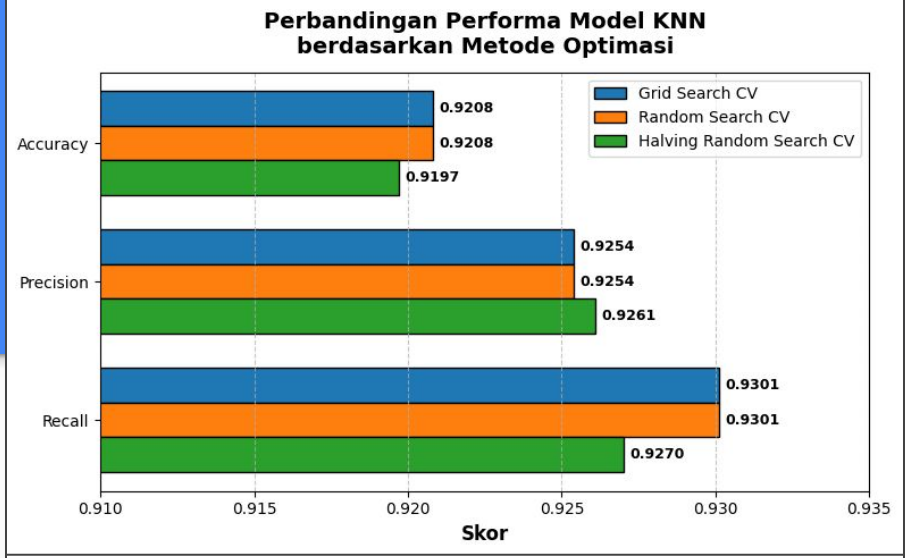
- Random Search is far more efficient than Grid Search in large search spaces (Bergstra & Bengio, 2012).
- Successive Halving and Hyperband reduce evaluation time significantly (Li et al., 2018).

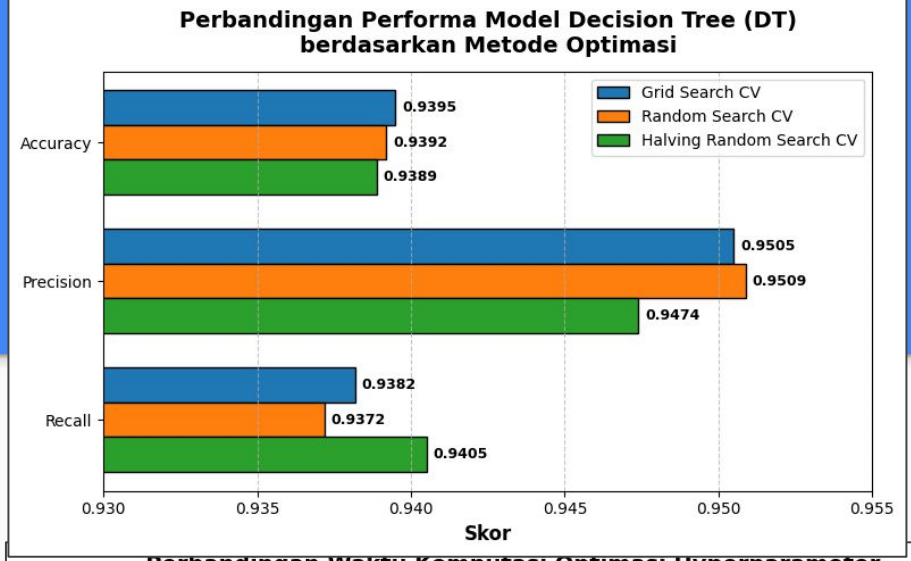
Novelty & Objective

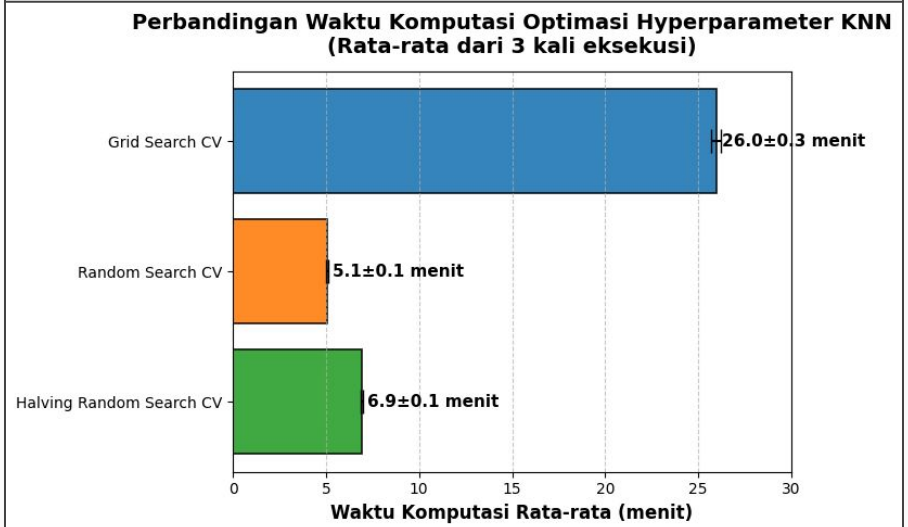
- Comparative analysis of hyperparameter optimization methods applied to four algorithms: KNN, Decision Tree, SVM, and Gaussian Naive Bayes.
- Focus: Balance between computational time efficiency and model quality (accuracy, precision, recall).
- Dataset : https://www.kaggle.com/datasets/mehmetsabrikunt/internet-service-churn

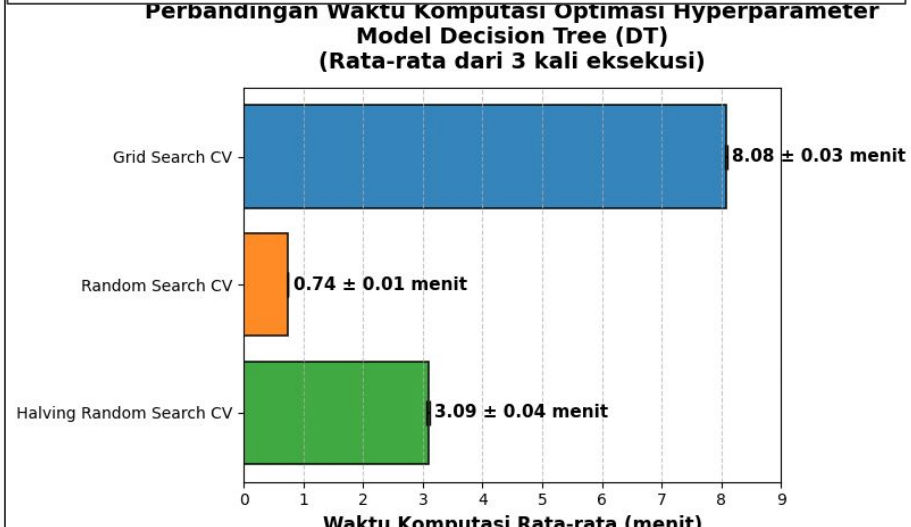
Methodology

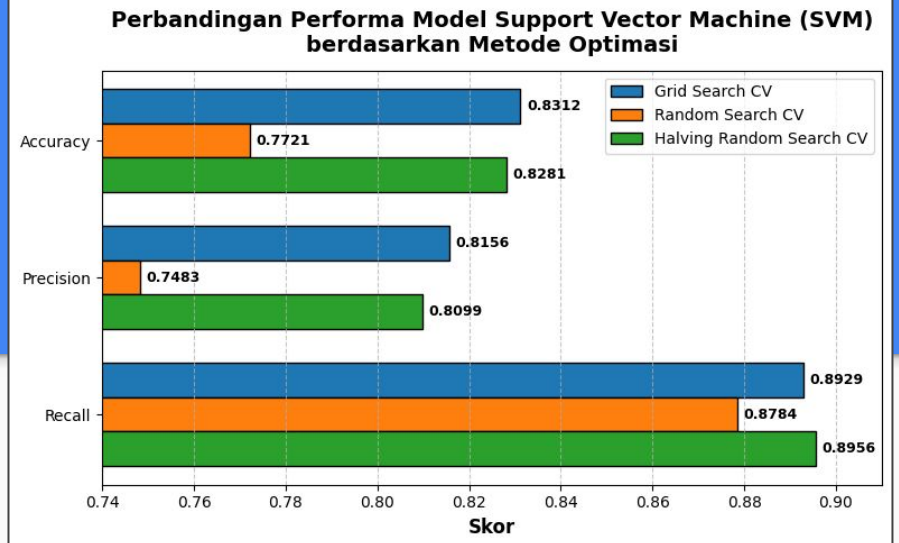
- Dataset: Internet Service Churn (72,274 samples, 10 features, 1 label; balanced class: 55.4% churn, 44.6% non-churn).
- Preprocessing: Median/mean imputation for missing values, remove irrelevant ID feature, Min-Max Scaling (0–1), 90% train – 10% test split.
- Models: KNN, Decision Tree, SVM, Gaussian Naive Bayes.
- Evaluation: 10-fold CV, repeated 3 times; metrics: accuracy, precision, recall; time measured with perf_counter.

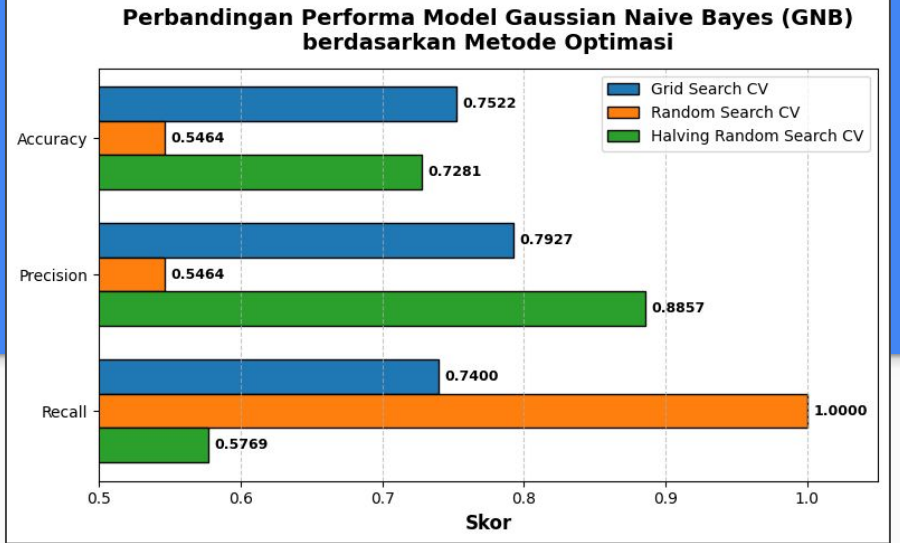

Perbandingan Performa Model KNN
berdasarkan Metode Optimasi
Grid Search CV
Random Search CV
Halving Random Search CV
Accuracy
Precision
Recall
0.9208
0.9208
0.9197
0.9254
0.9254
0.9261
0.9301
0.9301
0.9270
0.910
0.915
0.920
0.925
0.930
0.935
Skor

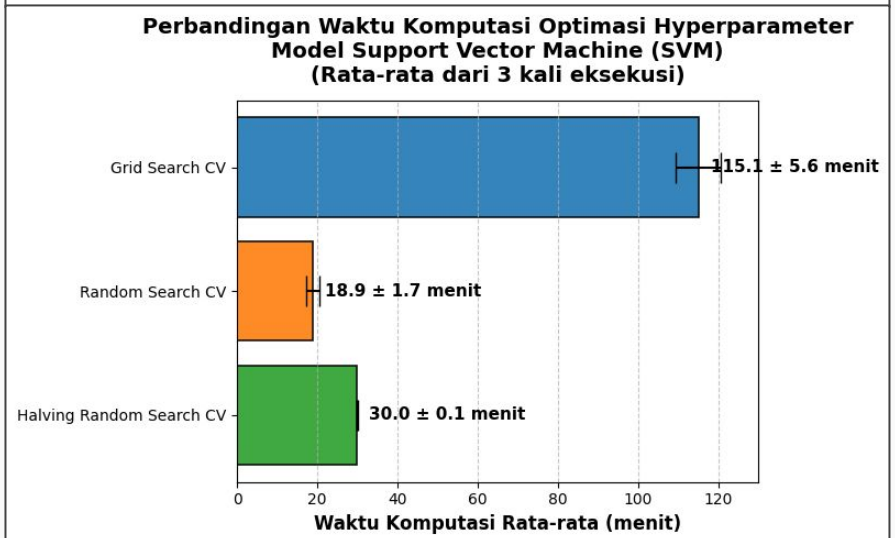

Perbandingan Performa Model Decision Tree (DT)
berdasarkan Metode Optimasi
Grid Search CV
Random Search CV
Halving Random Search CV
Accuracy
Precision
Recall
0.9395
0.9392
0.9389
0.9505
0.9509
0.9474
0.9382
0.9372
0.9405
0.930
0.935
0.940
0.945
0.950
0.955
Skor

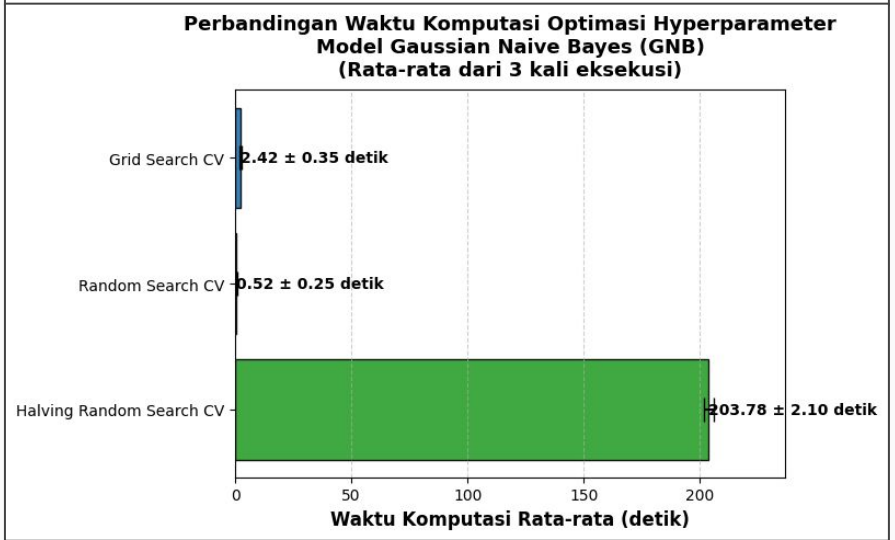

Perbandingan Waktu Komputasi Optimasi Hyperparameter KNN
(Rata-rata dari 3 kali eksekusi)
Grid Search CV
Random Search CV
Halving Random Search CV
26.0±0.3 menit
5.1±0.1 menit
6.9±0.1 menit
0
5
10
15
20
25
30
Waktu Komputasi Rata-rata (menit)


Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Decision Tree (DT)
(Rata-rata dari 3 kali eksekusi)
Grid Search CV
Random Search CV
Halving Random Search CV
8.08 ± 0.03 menit
0.74 ± 0.01 menit
3.09 ± 0.04 menit
0
1
2
3
4
5
6
7
8
9
Waktu Komputasi Rata-rata (menit)


Perbandingan Performa Model Support Vector Machine (SVM)
berdasarkan Metode Optimasi
Grid Search CV
Random Search CV
Halving Random Search CV
Accuracy
Precision
Recall
0.8312
0.7721
0.8281
0.8156
0.7483
0.8099
0.8929
0.8784
0.8956
0.74
0.76
0.78
0.80
0.82
0.84
0.86
0.88
0.90
Skor


Perbandingan Performa Model Gaussian Naive Bayes (GNB)
berdasarkan Metode Optimasi
Grid Search CV
Random Search CV
Halving Random Search CV
Accuracy
Precision
Recall
0.7522
0.5464
0.7281
0.7927
0.5464
0.8857
0.7400
1.0000
0.5769
0.5
0.6
0.7
0.8
0.9
1.0
Skor


Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Support Vector Machine (SVM)
(Rata-rata dari 3 kali eksekusi)
Grid Search CV
Random Search CV
Halving Random Search CV
115.1 ± 5.6 menit
18.9 ± 1.7 menit
30.0 ± 0.1 menit
0
20
40
60
80
100
120
Waktu Komputasi Rata-rata (menit)


Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Gaussian Naive Bayes (GNB)
(Rata-rata dari 3 kali eksekusi)
Grid Search CV
Random Search CV
Halving Random Search CV
2.42 ± 0.35 detik
0.52 ± 0.25 detik
203.78 ± 2.10 detik
0
50
100
150
200
Waktu Komputasi Rata-rata (detik)

Conclusion & Recommendation

- Halving achieves competitive performance with 62–74% time savings.
- Random Search is fastest but less stable.
- Halving slower in small hyperparameter space (overhead).

Future work: Integrate Hyperband or Bayesian Optimization for further improvements. Extend comparison to other algorithms or datasets.

Thank You!

Alief Cahyo Utomo
aliepcahyo@gmail.com
https://github.com/nixon42/

ANY QUESTION ?

Source Code :
https://github.com/nixon42/analisis_halving_random_search