

ANALISIS HALVING RANDOM SEARCH CROSS VALIDATION PADA OPTIMASI MODEL MACHINE LEARNING

Alief Cahyo Utomo

Universitas Nusantara PGRI Kediri, Jl. Ahmad Dahlan No.76, Mojoroto, Kec. Mojoroto, Kota Kediri, Jawa
Timur, 64112, Indonesia

e-mail: aliepcahyo@gmail.com

Abstract: Penelitian ini bertujuan untuk menganalisis efektivitas metode Halving Random Search Cross Validation sebagai alternatif optimasi hyperparameter pada model machine learning dibandingkan dengan Grid Search Cross Validation dan Random Search Cross Validation. Dataset yang digunakan adalah Internet Service Churn dengan empat algoritma: KNN, Decision Tree, SVM, dan Gaussian Naive Bayes. Proses pengujian melibatkan 10-fold cross validation dan pengulangan tiga kali untuk memastikan validitas hasil.

Hasil eksperimen menunjukkan bahwa Halving Random Search Cross Validation mampu mencapai performa akurasi, presisi, dan recall yang kompetitif (selisih $< 0,5\%$) dibandingkan Grid Search pada sebagian besar model, dengan penghematan waktu komputasi hingga 62–74% pada KNN, Decision Tree, dan SVM. Namun, pada Gaussian Naive Bayes dengan ruang hyperparameter kecil, metode ini lebih lambat karena overhead successive halving. Random Search menunjukkan kecepatan tinggi tetapi kurang stabil pada SVM dan Gaussian Naive Bayes.

Kesimpulan penelitian menyatakan bahwa Halving Random Search Cross Validation adalah metode paling seimbang untuk kasus bisnis seperti prediksi churn, dengan rekomendasi penerapan pada model kompleks dan pengembangan lanjutan menggunakan Hyperband atau Bayesian Optimization.

Key Words: Halving Random Search Cross-Validation, Hyperparameter Optimization, Machine Learning

Introduction

Di era digital saat ini, machine learning digunakan berbagai aplikasi industri mulai dari kesehatan, keuangan, hingga telekomunikasi. Kinerja model sangat ditentukan oleh konfigurasi hyperparameter yang optimal. Namun, proses hyperparameter tuning pada dataset berskala besar sering kali memakan waktu dan sumber daya komputasi yang sangat besar, terutama ketika menggunakan metode tradisional seperti Grid Search Cross Validation yang mengevaluasi seluruh kombinasi parameter.

Beberapa penelitian sebelumnya telah membandingkan metode optimasi hyperparameter. Bergstra dan Bengio (2012) menunjukkan bahwa Random Search Cross Validation jauh lebih efisien daripada Grid Search pada ruang pencarian besar. Li et al. (2018) kemudian memperkenalkan Successive Halving dan Hyperband sebagai pendekatan berbasis resource allocation yang mampu mengurangi waktu evaluasi model secara drastis. scikit-learn sejak versi 0.24 telah mengimplementasikan Halving Random Search Cross Validation sebagai kombinasi keduanya, namun studi empiris yang membandingkan performa dan efisiensi Halving Random Search Cross Validation secara langsung dengan Grid Search dan Random Search pada berbagai algoritma machine learning masih sangat terbatas.

Kebaruan penelitian ini terletak pada analisis komparatif terhadap Halving Random Search Cross Validation dibandingkan Grid Search dan Random Search menggunakan empat algoritma yang berbeda (KNN, Decision Tree, SVM, Gaussian Naive Bayes) pada dataset churn telekomunikasi (72.274 baris), dengan fokus ganda pada efisiensi waktu komputasi dan performa metrik (akurasi, presisi, recall).

Tujuan penelitian ini adalah untuk menganalisis dan membuktikan bahwa Halving Random Search Cross Validation mampu memberikan keseimbangan optimal antara efisiensi waktu komputasi dan kualitas model pada kasus prediksi churn pelanggan, sehingga dapat direkomendasikan sebagai metode standar hyperparameter tuning di lingkungan industri dengan keterbatasan sumber daya komputasi.

Method

Prosedur pengumpulan data mengikuti kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai acuan sistematis (Nila et al., 2023), dimulai dari tahap Business Understanding yang fokus pada evaluasi efektivitas Halving Random Search Cross Validation dibandingkan metode lain, dilanjutkan dengan Data Understanding untuk memahami dataset churn, dan Data Preparation melalui penghapusan data kosong serta normalisasi menggunakan Min-Max Scaler. Pada tahap Modeling, dilakukan optimasi hyperparameter terhadap empat algoritma (KNN, Decision Tree, SVM, dan Gaussian Naive Bayes) dengan 10-fold cross validation, di mana Grid Search dan Random Search menggunakan rentang hyperparameter diskrit seperti pada Tabel 1 dengan total 456 kombinasi untuk Grid dan 46 iterasi untuk Random setara 10% dari total, sementara Halving menggunakan distribusi kontinu seperti loguniform dan randint pada Tabel 2 untuk ruang pencarian yang lebih luas. Setiap eksperimen diulang tiga kali untuk mendapatkan nilai rata-rata.

Tabel 1. List Hyperparameter 1

Model	Hyperparameter	Nilai	Jumlah kombinasi unik
KNN	n_neighbor	3, 5, 7, 9, 11, 13, 15, 17, 19, 21	200 Kombinasi
	algorithm	'ball_tree', 'kd_tree'	
	weight	uniform, distance	
	metric	'cityblock', 'cosine', 'euclidean', 'manhattan', 'nan_euclidean'	
Decision Tree	criterion	'gini', 'entropy', dan 'log_loss'	240 Kombinasi
	max_depth	None, 3, 5, 7, 8, 9, 10, 15, 20, 25	
	min_sample_leaf	10, 20, 30, 40, 50, 100, 200, 500	
SVM	C	0.001, 1.0	6 Kombinasi
	kernel	'linear', 'rbf', 'poly'	
Gaussian Naive Bayes	var_smoothing	0.001, 0.00316, 0.01, 0.0316, 0.1, 0.316, 1.0, 3.16, 10.0, 100.0	10 Kombinasi

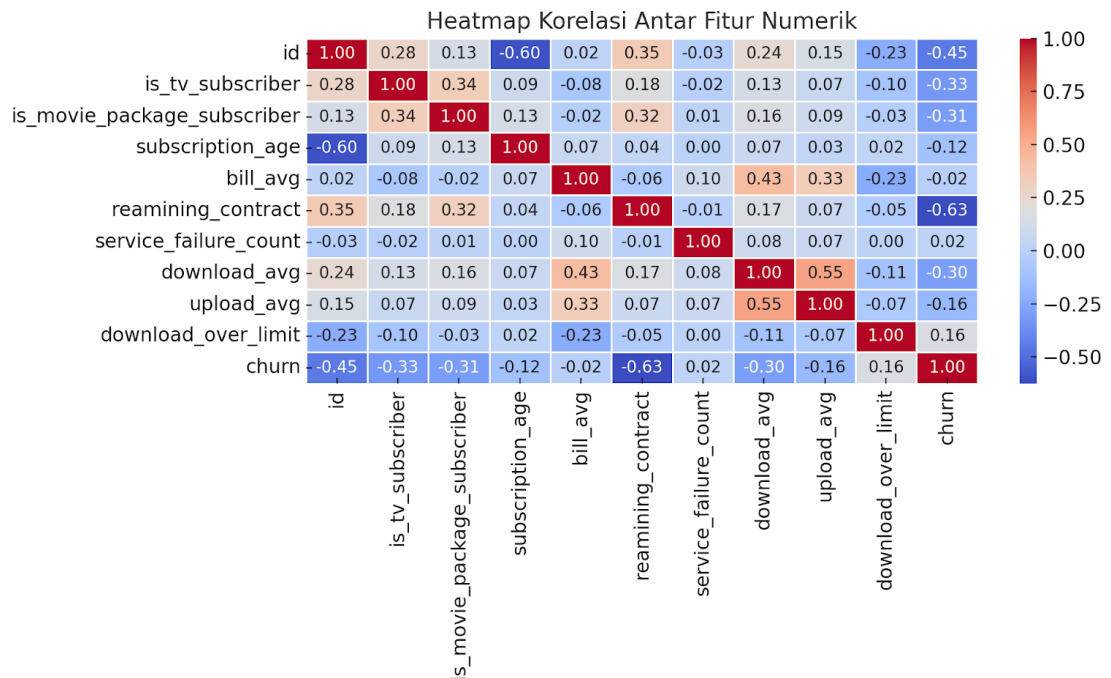
Tabel 2. List Hyperparameter 2

Model	Hyperparameter	Nilai	Jumlah kombinasi unik
KNN	n_neighbor	'randint(3, 30)'	810 Kombinasi
	algorithm	'ball_tree', 'kd_tree'	
	weight	uniform, distance	
	metric	'cityblock', 'cosine', 'euclidean', 'manhattan', 'nan_euclidean'	
Decision Tree	criterion	'gini', 'entropy', dan 'log_loss'	37,500 Kombinasi
	max_depth	'[None, range(3, 26)]'	
	min_sample_leaf	'randint(10, 501)'	
SVM	C	'loguniform(1e-3, 1e2)'	Tidak terbatas
	kernel	'linear', 'rbf', 'poly'	
Gaussian Naive Bayes	var_smoothing	'loguniform(1e-3, 1e2)'	Tidak terbatas

Results and Discussion

1. Hasil Explorasi Dataset

Hasil eksplorasi data menunjukkan bahwa dataset terdiri dari 72.274 data, 10 fitur dan 1 label dengan distribusi kelas yang relatif seimbang (55,4% churn dan 44,6% non-churn). Analisis korelasi memperlihatkan bahwa lama berlangganan dan sisa kontrak berkorelasi negatif terhadap churn. Beberapa fitur memiliki outlier, namun tetap dipertahankan karena merepresentasikan kondisi nyata pelanggan.

**Gambar 1.** Heatmap Korelasi Antar Fitur

2. Pra-Pemrosesan Dataset

Data dibersihkan dari nilai kosong dengan imputasi median pada remaining contract dan rata-rata pada download average serta upload average. Fitur id dihapus karena tidak relevan secara prediktif. Seluruh fitur numerik dinormalisasi menggunakan Min–Max Scaling ke rentang 0–1. Dataset kemudian dibagi menjadi 90% data latih dan 10% data uji.

3. Proses Pelatihan

Model yang dipilih adalah KNN, Decision Tree, SVM, dan Gaussian Naive Bayes. Optimasi hyperparameter dilakukan dengan tiga metode, yaitu Grid Search Cross Validation, Random Search Cross Validation, dan Halving Random Search Cross Validation menggunakan 10-fold cross validation. Setiap eksperimen diulang sebanyak tiga kali, dan waktu komputasi diukur menggunakan fungsi *perf_counter* untuk mengukur waktu yang diperlukan masing-masing metode.

Tabel 3. Metode Grid Search Cross Validation

Model	Run-ke	Waktu (detik)	Akurasi	Presisi	Recall
KNN	1	1540.61	0.9208	0.9254	0.9301
	2	1556.37			
	3	1577.01			
Decision Tree	1	487.07	0.9395	0.9505	0.9382
	2	484.42			
	3	483.17			
SVM	1	7319.16	0.8312	0.8156	0.8929
	2	6892.50			
	3	6501.26			
GNB	1	2.22	0.7522	0.7927	0.74
	2	2.91			
	3	2.13			

Tabel 4. Metode Random Search Cross Validation

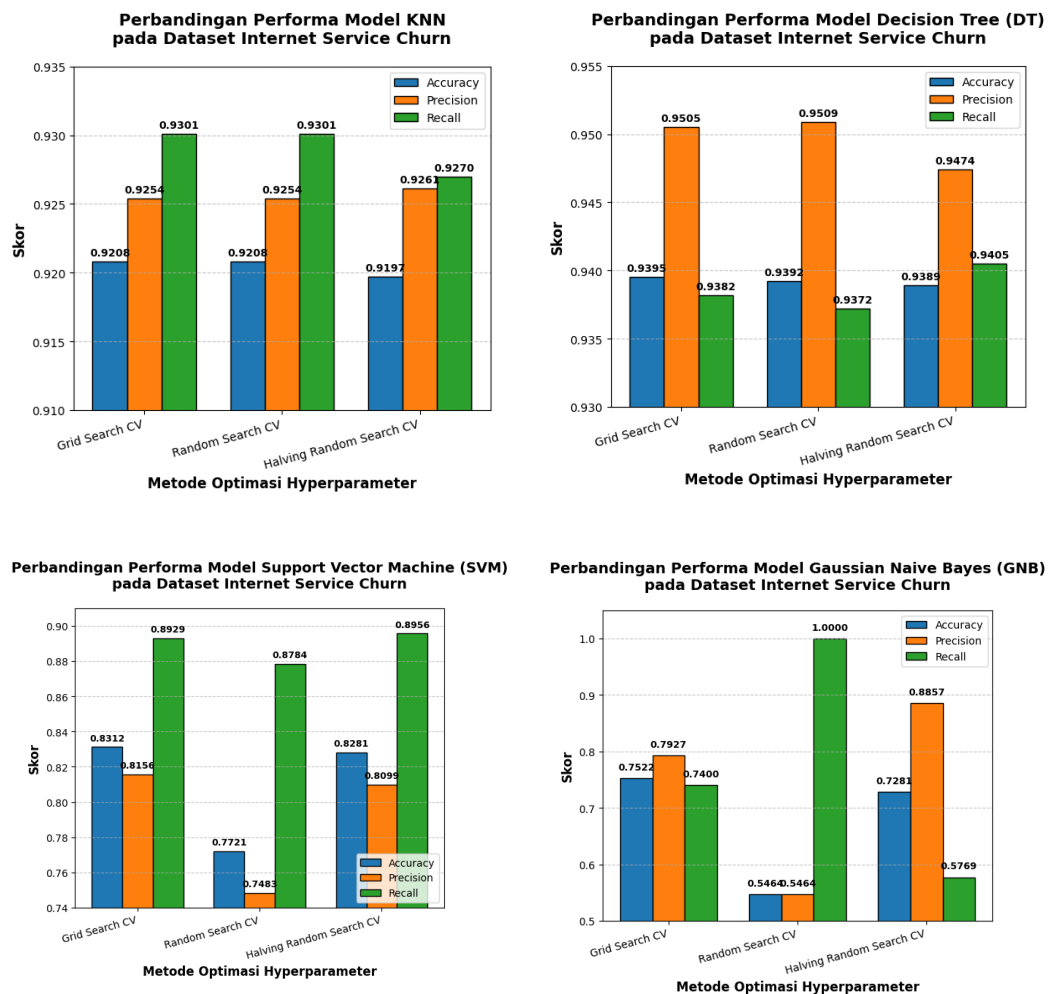
Model	Run-ke	Waktu (detik)	Akurasi	Presisi	Recall
KNN	1	301.37	0.9208	0.9254	0.9301
	2	299.22			
	3	309.18			
Decision Tree	1	44.67	0.9392	0.9509	0.9372
	2	43.86			
	3	44.56			
SVM	1	1121.51	0.7721	0.74832	0.8784
	2	1261.76			
	3	1014.56			
GNB	1	0.87	0.5464	0.5464	1.0
	2	0.35			
	3	0.34			

Tabel 5. Metode Halving Random Search Cross Validation

Model	Run-ke	Waktu (detik)	Akurasi	Presisi	Recall
KNN	1	419.95	0.9197	0.9261	0.9270
	2	412.06			
	3	411.08			
Decision Tree	1	182.32	0.9389	0.9474	0.9405
	2	187.44			
	3	186.31			
SVM	1	1804.39	0.8281	0.8099	0.8956
	2	1800.82			
	3	1794.14			
GNB	1	204.32	0.7281	0.8857	0.5769
	2	200.98			
	3	206.03			

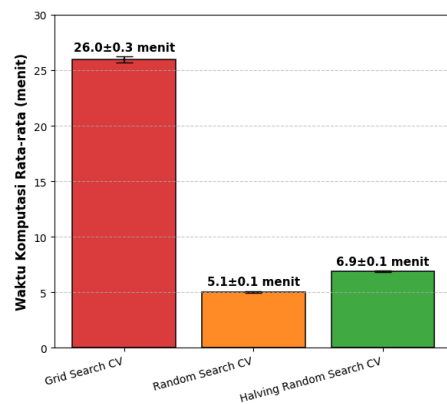
Conclusion

Berdasarkan hasil pengujian, Halving Random Search Cross Validation terbukti mampu memberikan keseimbangan yang optimal antara kualitas model dan efisiensi waktu komputasi dibandingkan metode Grid Search Cross Validation dan Random Search Cross Validation. Halving Random Search secara signifikan mengurangi waktu komputasi pada sebagian besar model, terutama pada SVM, Decision Tree, dan KNN. Meskipun pada Gaussian Naive Bayes metode ini menjadi relatif lebih lambat akibat overhead successive halving pada ruang hyperparameter yang sempit. Secara keseluruhan, metode ini mampu mencapai selisih akurasi kurang dari 0,5% dibandingkan Grid Search dan penghematan waktu komputasi sebesar 62–74%.

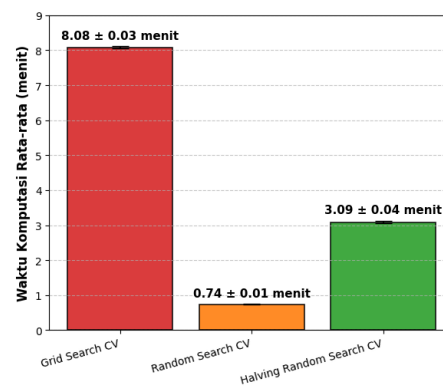


Gambar 2. Perbandingan Akurasi, Presisi, dan Recall

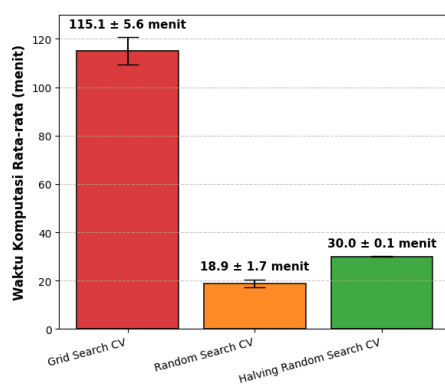
**Perbandingan Waktu Komputasi Optimasi Hyperparameter K
(Rata-rata dari 3 kali eksekusi)**



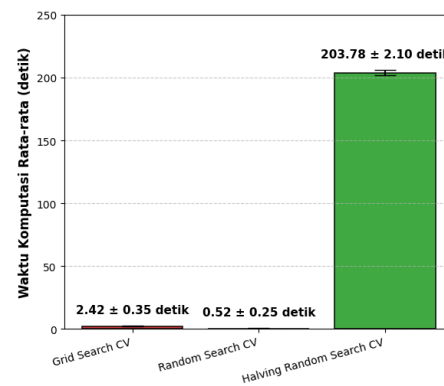
**Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Decision Tree (DT)
(Rata-rata dari 3 kali eksekusi)**



**Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Support Vector Machine (SVM)
(Rata-rata dari 3 kali eksekusi)**



**Perbandingan Waktu Komputasi Optimasi Hyperparameter
Model Gaussian Naive Bayes (GNB)
(Rata-rata dari 3 kali eksekusi)**



Gambar 3. Perbandingan Waktu

References (should use Mendeley program)

- Nurjanah, I., Karaman, J., Widaningrum, I., Mustikasari, D., & Sucipto. (2023). Penggunaan Algoritma Naive Bayes Untuk Menentukan Pemberian Kredit Pada Koperasi Desa. In Journal of Computer Science and Information Technology E-ISSN (Vol. 3, Issue 2). <https://doi.org/https://doi.org/10.47065/explorer.v3i2.766>
- Sucipto, Prasetya, D. D., & Widiyaningtyas, T. (2025). A Supervised Hybrid Weighting Scheme for Bloom's Taxonomy Questions using Category Space Density-based Weighting. Engineering, Technology and Applied Science Research, 15(2), 22102–22108. <https://doi.org/10.48084/etasr.10226>
- Nila, U., Firliana, R., & Sucipto. (2023). Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi) 175 Analisis Data Transaksi Penjualan Produk Pertanian Menggunakan Algoritma FP-Growth Kata Kunci-data transaksi, data mining, Algoritma FP-Growth. In Agustus (Vol. 7). <https://doi.org/10.29407/inotek.v7i1.3426>
- Anggraini, I. Y., Sucipto, S., & Indriati, R. (2018). Cyberbullying Detection Modelling at Twitter Social Networking. JUITA: Jurnal Informatika, 6(2), 113–118. <https://doi.org/10.30595/juita.v6i2.3350>
- Aprilliandhika, W., & Fauzi Abdulloh, F. (2024). COMPARISON OF K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE ALGORITHM OPTIMIZATION WITH GRID SEARCH CV ON STROKE PREDICTION KOMPARASI OPTIMASI ALGORITMA K-NEAREST NEIGHBOR DAN

- SUPPORT VECTOR MACHINE DENGAN GRID SEARCH CV PADA PREDIKSI STROKE. 5(4), 991–1000. <https://doi.org/10.52436/1.jutif.2024.5.4.1951>
- Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Irmanda, H. N., & Astriratma, R. (2017). Terakreditasi SINTA Peringkat 2 Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM). Masa Berlaku Mulai, 1(3), 915–922. <https://doi.org/10.29207/resti.v4i5.2313>
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18(185), 1–52.
- Bergstra, J., Ca, J. B., & Ca, Y. B. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. In *Journal of Machine Learning Research* (Vol. 13).
- Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Irmanda, H. N., & Astriratma, R. (2017). Terakreditasi SINTA Peringkat 2 Klasifikasi Jenis Pantun dengan Metode Support Vector Machines (SVM). Masa Berlaku Mulai, 1(3), 915–922. <https://doi.org/10.29207/resti.v4i5.2313>
- Irawan, I., Qisthiano, R., Syahril, M., & Jakak, P. M. (2023). Optimasi Prediksi Kelulusan Tepat Waktu: Studi Perbandingan Algoritma Random Forest dan Algoritma K-NN Berbasis PSO. In *Jurnal Pengembangan Sistem Informasi dan Informatika* (Vol. 4, Issue 4). <https://doi.org/10.47747/jpsii.v4i4.1374>
- Jamiluddin, F., Faisal, S., Arum Puspita Lestari, S., Fauzi, A., Karawang, P., Jl HSRonggo Waluyo, K., Timur, T., & Barat, J. (2024). Implementasi Hyperparameter Tuning Grid Search CV Pada Prediksi Produksi Padi Menggunakan Algoritma Linear Regresi. *Journal of Information System Research*, 6(1), 490–498. <https://doi.org/10.47065/josh.v6i1.5930>
- Misnawati. (2023). ChatGPT: Keuntungan, Risiko, Dan Penggunaan Bijak Dalam Era Kecerdasan Buatan. In *Jurnal Prosiding Mateandrau* (Vol. 2, Issue 1). <https://doi.org/10.55606/mateandrau.v2i1.221>
- Muhamad, I., & Matin, M. (2023). Hyperparameter Tuning menggunakan GridsearchCV pada Random Forest untuk Deteksi Malware. <https://doi.org/10.32722/multinetics.v9i1.5578>
- Munawaroh, S., Rosyidah, U. A., & Yanuarti, R. (2024). Klasifikasi Tingkat Kecemasan Atlet Sebelum Bertanding Menggunakan Algoritma K-Nearest Neighbor (KNN) Berbasis Website. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 5(2), 87–94. <https://doi.org/10.37148/bios.v5i2.120>
- Nugraha, W., & Sasongko, A. (2022). SISTEMASI: Jurnal Sistem Informasi Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search Hyperparameter Tuning on Classification Algorithm with Grid Search (Vol. 11, Issue 2). <https://doi.org/10.32520/stmsi.v11i2.1750>
- Nugroho, A., Soeleman, Ma., Anggi Pramunendar, R., & Nurhindarto, A. (2023). PENINGKATAN PERFORMA ENSEMBLE LEARNING PADA SEGMENTASI SEMANTIK GAMBAR DENGAN TEKNIK OVERSAMPLING UNTUK CLASS IMBALANCE. <https://doi.org/10.25126/jtiik.2023106831>
- Putri, T. A. E., Widiharih, T., & Santoso, R. (2023). PENERAPAN TUNING HYPERPARAMETER RANDOMSEARCHCV PADA ADAPTIVE BOOSTING UNTUK PREDIKSI KELANGSUNGAN

- HIDUP PASIEN GAGAL JANTUNG. *Jurnal Gaussian*, 11(3), 397–406. <https://doi.org/10.14710/j.gauss.11.3.397-406>
- Rahmat, A., Syafiih, M., & Faid, M. (2023). IMPLEMENTASI KLASIFIKASI POTENSI PENYAKIT JANTUNG DENGAN MENGGUNAKAN METODE C4.5 BERBASIS WEBSITE (STUDI KASUS KAGGLE.COM). *INFOTECH Journal*, 9(2), 393–400. <https://doi.org/10.31949/infotech.v9i2.6295>
- Sartika, D., & Sensuse, D. I. (2017). Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian (Vol. 1, Issue 2). <https://doi.org/10.35957/jatisi.v3i2.78>
- Soper, D. S. (2023). Hyperparameter Optimization Using Successive Halving with Greedy Cross Validation. *Algorithms*, 16(1). <https://doi.org/10.3390/a16010017>