

## ANALISIS KOMPARATIF METODE HYPERPARAMETER TUNING PADA MODEL KLASIFIKASI UNTUK DATA BALANCED DAN INBALANCED

Anas Syaifudin<sup>1)</sup>, Indrayanti<sup>2)</sup>, Hermanus Wim Hapsoro<sup>3)</sup>, Rizqi Wijonarko<sup>4)</sup>

Institut Widya Pratama<sup>123)</sup>

anas.sya@iwima.ac.id<sup>1)</sup>, indrayanti3214@gmail.com<sup>2)</sup>, wimhapsoro06@gmail.com<sup>3)</sup>,  
wjonarkorizqi@gmail.com<sup>4)</sup>

### Abstrak

Pemilihan metode hyperparameter tuning yang tepat berperan penting dalam meningkatkan kinerja model klasifikasi, terutama ketika diterapkan pada dataset dengan karakteristik distribusi kelas yang berbeda. Penelitian ini bertujuan untuk menganalisis dan membandingkan efektivitas tiga metode hyperparameter tuning, yaitu Grid Search, Random Search, dan Bayesian Optimization, pada model XGBoost, Random Forest, dan Support Vector Machine (SVM). Pengujian dilakukan menggunakan dua dataset dengan karakteristik berbeda, yaitu Breast Cancer Wisconsin sebagai dataset berimbang dan Credit Card Fraud Detection sebagai dataset tidak berimbang. Evaluasi kinerja model disesuaikan dengan karakteristik dataset, menggunakan F1-score (macro) untuk dataset Breast Cancer dan Precision-Recall AUC untuk dataset Credit Card Fraud. Hasil penelitian menunjukkan bahwa pada dataset berimbang, seluruh metode tuning menghasilkan performa yang relatif serupa, dengan SVM secara konsisten memberikan hasil terbaik. Sebaliknya, pada dataset tidak berimbang, metode Random Search dan Bayesian Optimization menunjukkan keunggulan dalam menemukan konfigurasi hyperparameter yang mampu meningkatkan kemampuan deteksi kelas minoritas, khususnya pada model XGBoost. Temuan ini menegaskan bahwa pemilihan metode tuning dan metrik evaluasi harus disesuaikan dengan karakteristik data yang digunakan.

**Kata kunci:** Hyperparameter Tuning, XGBoost, Random Forest, Support Vector Machine, Dataset Tidak Berimbang.

### 1. Pendahuluan

Kecerdasan buatan (Artificial Intelligence/AI) dan pembelajaran mesin (Machine Learning/ML) telah menjadi penggerak utama transformasi digital di berbagai sektor, mulai dari kesehatan, keuangan, hingga e-commerce (Bishop et al. 2023). Kemampuan ML untuk mempelajari pola dari data dan membuat prediksi yang akurat telah menjadikannya tulang punggung dalam pengambilan keputusan berbasis data. Dalam konteks pembelajaran terawasi (supervised learning), klasifikasi merupakan salah satu tugas yang paling banyak diaplikasikan. Keberhasilan sebuah model klasifikasi tidak hanya bergantung pada pemilihan algoritma yang tepat, tetapi juga pada optimalisasi hyperparameter yang mendasarinya (Bischl et al. 2023). Proses penyesuaian hyperparameter, yang dikenal sebagai hyperparameter tuning, adalah langkah penting untuk memaksimalkan akurasi, generalisasi, dan efisiensi model.

Hyperparameter, yang nilainya ditetapkan sebelum proses pelatihan, seperti tingkat pembelajaran (learning rate) pada XGBoost atau parameter C dan gamma pada Support Vector Machine (SVM), memainkan peran penting dalam menentukan performa akhir model.

Tetapi pengaturan hyperparameter menghadapi banyak masalah metodologis yang signifikan. Metode tradisional seperti Grid Search, meskipun menjamin menemukan kombinasi terbaik dalam ruang pencarian yang telah ditentukan, memiliki kelemahan dalam hal kompleksitas komputasi yang meningkat (Michalakopoulos et al. 2025). Dataset besar atau model yang kompleks tidak dapat menggunakan metode ini. Random Search, yang memilih kombinasi hyperparameter secara acak, menunjukkan efisiensi yang lebih baik dan sering kali menemukan solusi yang memadai dengan lebih cepat (Tiastama and Budi 2024). Namun, pendekatan ini tetap tidak memiliki mekanisme

yang baik untuk mempelajari dari evaluasi sebelumnya.

Sebagai respon atas keterbatasan metode tradisional, teknik canggih seperti Bayesian Optimization hadir dengan pendekatan yang lebih adaptif dan efisien. Teknik ini membangun model probabilistik dari fungsi objektif (seperti skor validasi silang) dan menggunakannya untuk melakukan pencarian ke area yang paling potensial dalam ruang hyperparameter (Morales-Hernández, Nieuwenhuys, and Gonzalez 2022). Bayesian Optimization dapat menemukan konfigurasi optimal dengan jumlah evaluasi yang jauh lebih sedikit dengan menggunakan data dari iterasi sebelumnya. Ini menghasilkan efisiensi komputasional yang lebih tinggi. Namun, kompleksitas implementasinya dan kebutuhan untuk memahami dasar teorinya sering kali menjadi penghalang (Bischi et al. 2023).

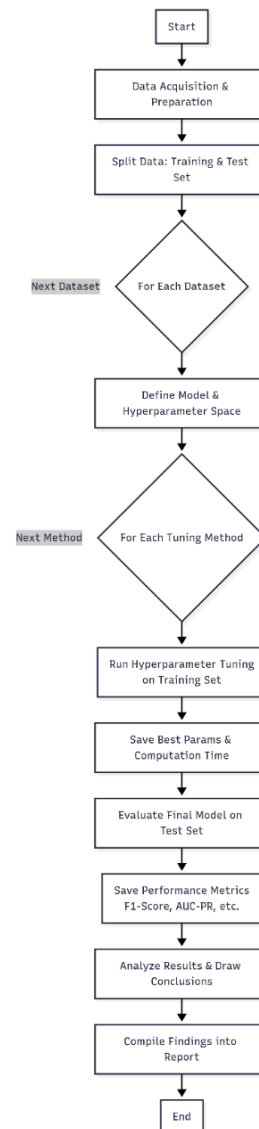
Permasalahan menjadi semakin kompleks ketika model klasifikasi dihadapkan pada dataset dengan distribusi kelas yang tidak seimbang (imbalanced dataset). Permasalahan pada data yang tidak seimbang sering ditemui, seperti pada deteksi penipuan transaksi keuangan atau diagnosis penyakit langka (Kaur, Pannu, and Malhi 2019). Dalam kasus ketidakseimbangan kelas, metrik evaluasi konvensional seperti akurasi menjadi sangat membingungkan. Ini karena model hanya dapat memprediksi kelas mayoritas dengan akurasi tinggi (Johnson and Khoshgoftaar 2019). Oleh karena itu, hyperparameter tuning dalam konteks ini tidak boleh berfokus pada akurasi, tetapi harus mengoptimalkan metrik, seperti skor F1-Score atau Area Under the Precision-Recall Curve (AUC-PR) (Parekh, Nemlekar, and Losey 2025).

Banyak penelitian sebelumnya yang telah berfokus pada metode tuning secara terisolir atau pada model tertentu. Namun, belum ada penelitian komparatif yang menyeluruh dan terbaru yang membandingkan ketiga metode tuning (Grid Search, Random Search, dan Bayesian Optimization) secara bersamaan pada model klasifikasi populer seperti XGBoost, Random Forest, dan Support Vector Machine (SVM) untuk kedua jenis dataset (berimbang dan tidak berimbang) dengan menggunakan metrik yang sesuai untuk setiap situasi (Dasgupta and Sen 2024). Dengan memahami kelebihan dan

kekurangan setiap metode tuning pada berbagai skenario dataset, selanjutnya dapat mengalokasikan sumber daya komputasional secara lebih efektif dan efisien.

## 2. Metode Penelitian

Dalam melakukan penelitian ini disusun rancangan atau kerangka pemikiran untuk sebagai acuan dalam menyelesaikan penelitian, agar sesuai dengan pokok bahasan yang diteliti. Pada Gambar 1 Berikut ini menunjukkan kerangka pemikiran yang digunakan dalam penelitian ini.



Gambar 1. Alur Penelitian

Metode penelitian menjelaskan tahapan penelitian atau pengembangan yang dilakukan

untuk mencapai tujuan/sasaran penelitian. Tiap tahap dijelaskan secara ringkas, misalnya tiap tahap dalam satu paragraf. Bahan/materi/platform yang digunakan dalam penelitian diuraikan di bab ini, yaitu meliputi subjek/bahan yang diteliti, alat / perangkat lunak bantu yang digunakan, rancangan percobaan atau desain yang digunakan, teknik pengambilan sampel, rencana pengujian (variabel yang akan diukur dan teknik mengambil data), analisis dan model statistik yang digunakan.

## 2.1 Akuisisi dan Persiapan Data

Download dataset Breast Cancer (balanced) dan Credit Card Fraud (imbalanced). Lakukan preprocessing seperti handling missing values, feature scaling, dan encoding.

## 2.2 Pembagian Data

Pisahkan setiap dataset menjadi training set (80%) dan test set (20%). Gunakan stratified sampling untuk menjaga distribusi kelas.

## 2.3 Definisi Model dan Ruang Pencarian

Untuk setiap model (XGBoost, RF, SVM), definisikan ruang hyperparameter yang akan di-tuning.

## 2.4 Pelaksanaan Tuning

Untuk setiap metode tuning (Grid, Random, Bayesian), lakukan pencarian hyperparameter terbaik pada training set menggunakan cross-validation.

## 2.5 Penyimpanan Hasil Tuning

Mencatat kombinasi hyperparameter terbaik dan waktu komputasi yang dibutuhkan oleh setiap metode.

## 2.6 Evaluasi Model Akhir

Latih ulang model pada seluruh training set menggunakan hyperparameter terbaik. Evaluasi performa model akhir yang telah dilatih ulang tersebut pada test set yang tidak pernah disentuh selama proses tuning. Catat metrik F1-Score, AUC-PR, dan Akurasi.

## 2.7 Evaluasi Model Akhir

Bandingkan hasil dari semua kombinasi (model x dataset x metode tuning) berdasarkan metrik

performa dan waktu komputasi. Tarik kesimpulan dan buat rekomendasi.

## 3. Hasil dan Pembahasan

Penelitian ini mengevaluasi kinerja tiga model klasifikasi, yaitu XGBoost, Random Forest, dan Support Vector Machine (SVM), yang dioptimasi menggunakan tiga metode hyperparameter tuning, yakni Grid Search, Random Search, dan Bayesian Optimization, pada dua dataset dengan karakteristik berbeda, yaitu dataset Breast Cancer Wisconsin (berimbang) dan Credit Card Fraud Detection (tidak berimbang).

### 3.1 Akuisisi dan Persiapan Data

Tahap awal pada penelitian difokuskan pada pengolahan dua karakteristik dataset yang berbeda guna menguji ketangguhan metode tuning. Dataset pertama adalah Breast Cancer Wisconsin yang merepresentasikan data cenderung seimbang (balanced), sedangkan dataset kedua adalah Credit Card Fraud Detection yang memiliki tingkat ketimpangan kelas (imbalanced) yang sangat ekstrem. Pada tahap ini, dilakukan proses data cleaning untuk memastikan integritas data. Mengingat algoritma Support Vector Machine (SVM) sangat bergantung pada perhitungan jarak antar data, dilakukan standarisasi menggunakan StandardScaler. Transformasi ini krusial agar fitur-fitur dengan skala besar tidak mendominasi proses pembelajaran model, sehingga setiap fitur memberikan kontribusi yang proporsional dalam penentuan batas keputusan (decision boundary).

Dataset Breast Cancer Wisconsin dengan jumlah fitur sebanyak 30 dan jumlah baris sebanyak 569 dengan distribusi label 357 data positif dan 212 data negatif. Dapat dilihat bahwa terdapat perbedaan rentang nilai pada setiap fiturnya sehingga akan dilakukan proses transformasi. Pada dataset Credit Card Fraud terdapat 31 fitur dan 284807 baris data dengan distribusi label 201 data positif dan 83077 data negatif. Perbedaan label yang terlalu jauh ini dapat dinyatakan bahwa dataset ini memiliki tingkat ketimpangan kelas yang tinggi. Pada dataset ini juga dilakukan transformasi dengan menggunakan StandardScaler.

Tabel 1. Dataset Breast Cancer Wisconsin

Nomor	Nama	Pekerjaan	Status
1	Ali	PNS	Menikah
2	Hasan	Swasta	Duda
3	Jonny	Petani	Janda
4	Rizal	Buruh	Lajang

Usahakan tabel jangan terpotong pada halaman yang berbeda, kecuali jika besarnya melebihi satu halaman. Jika harus terpotong, jangan lupa tulis ulang *header row* untuk setiap kolomnya, diberi nomor urut tabel yang sama, dan judul diganti dengan *Lanjutan*. Judul tabel tidak diakhiri dengan titik. Tabel tidak perlu menggunakan garis vertikal.

### 3.2 Pembagian Data

Tahap selanjutnya, dilakukan pembagian kedua dataset menjadi 70 % data latih dan 30 % data uji. Pembagian ini dilakukan dengan tujuan untuk mengevaluasi kinerja model machine learning yang akan digunakan secara objektif serta untuk memastikan bahwa model tersebut dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

Tabel 2. Distribusi pembagian dataset Breast Cancer dan Credit Card Fraud

Breast Cancer		Credit Card Fraud	
Train	Test	Train	Test
398	171	58294	24984

### 3.3 Definisi Model dan Ruang Pencarian

Dalam konteks pembelajaran mesin, khususnya saat melakukan hyperparameter tuning, definisi model merujuk pada pemilihan algoritma dasar yang akan digunakan untuk membangun sistem prediktif. Model yang akan digunakan adalah XGBoost, Random Forest, Support Vector Machine. Setiap model memiliki serangkaian hyperparameter internal yang tidak dipelajari langsung dari data, namun sangat memengaruhi performa dan kemampuan generalisasi model tersebut. Dari ketiga model yang dipilih nantinya akan dilakukan perbandingan. Berikut ini penjelasan tentang hyperparameter internal yang digunakan pada model:

#### a. Parameter model XGBoost

Secara default, XGBoost menggunakan beberapa parameter kunci seperti `objective = 'binary:logistic'` untuk klasifikasi biner, `eval_metric = 'logloss'`, `learning_rate (eta)` sebesar 0.3, `max_depth = 6`, serta jumlah estimator (`n_estimators`) 100. Parameter ini dirancang untuk memberikan kinerja yang baik pada berbagai kasus umum.

#### b. Parameter model Random Forest

Untuk Random Forest, parameter default meliputi `n_estimators=100` yang berarti membangun 100 pohon keputusan, `criterion = 'gini'` sebagai fungsi pengukuran kualitas pemisahan, `max_depth = None` yang memungkinkan pohon tumbuh tanpa batas hingga daun menjadi murni, `min_samples_split = 2`, `min_samples_leaf = 1`, dan `max_features = 'sqrt'` mempertimbangkan akar kuadrat dari jumlah fitur saat mencari pemisahan terbaik.

#### c. Parameter model Support Vector Machine

Model SVC memiliki parameter default seperti `C=1.0` (parameter regularisasi), `kernel='rbf'` (Radial Basis Function) sebagai fungsi kernel, dan `gamma='scale'` untuk koefisien kernel. `Probability = False` secara default, tetapi dalam kasus ini telah diatur ke `True` sebelumnya di notebook untuk memungkinkan perhitungan probabilitas.

### 3.4 Pelaksanaan Tuning

Ruang pencarian (hyperparameter search space) adalah rentang nilai atau kombinasi nilai yang telah ditentukan sebelumnya untuk setiap hyperparameter dari ketiga model yang akan digunakan. Ruang pencarian ini mendefinisikan batasan eksplorasi selama proses tuning, di mana algoritma optimasi akan mencari konfigurasi hyperparameter terbaik yang menghasilkan performa model optimal. Algoritma optimasi yang akan digunakan untuk mengoptimalkan parameter adalah Grid Search, Random Search dan Bayesian Optimization. Penentuan ruang pencarian yang efektif sangat krusial, karena

ruang yang terlalu sempit mungkin melewatkan konfigurasi yang optimal. Sementara ruang yang terlalu lebar dapat meningkatkan kompleksitas komputasi secara signifikan. Berdasarkan hal ini ketiga algoritma optimasi ini akan digunakan untuk melihat algoritma mana yang paling optimal dalam mengoptimalkan ruang pencarian pada dataset berimbang dan tidak berimbang. Berikut penjelasan detail konfigurasi ruang pencarian yang dilakukan untuk setiap algoritma optimasi:

a. Ruang Pencarian Hyperparameter pada Dataset Breast Cancer

Untuk dataset Breast Cancer, ruang pencarian hyperparameter dirancang relatif lebih luas karena ukuran dataset yang lebih kecil dan memungkinkan eksplorasi parameter yang lebih luas tetapi tanpa menggunakan banyak daya komputasi. Pada algoritma XGBoost, beberapa hyperparameter utama yang dieksplorasi meliputi jumlah pohon (`n_estimators`), learning rate, kedalaman maksimum pohon (`max_depth`), dan proporsi fitur yang digunakan dalam setiap pohon (`colsample_bytree`). Pada algoritma Random Forest, hyperparameter yang dioptimasi mencakup jumlah pohon, kedalaman maksimum pohon, jumlah minimum sampel untuk melakukan pemisahan node (`min_samples_split`), serta jumlah minimum sampel pada node daun (`min_samples_leaf`). Untuk algoritma Support Vector Machine (SVM) pada dataset Breast Cancer, ruang pencarian difokuskan pada parameter `C`, jenis kernel, dan parameter `gamma`. Pengaturan ruang pencarian yang digunakan untuk dataset Breast Cancer dapat dilihat pada Tabel 2 yang menjelaskan parameter metode tuning dan rentang nilai.

Tabel 2. Ruang Pencarian Hyperparameter pada Dataset Breast Cancer

Algoritma	Metode Tuning	Parameter	Rentang Nilai	
XGBoost	Grid Search	<code>n_estimators</code>	[100, 200, 300]	
		<code>learning_rate</code>	[0.01, 0.1, 0.2]	
		<code>max_depth</code>	[3, 5, 7]	
		<code>colsample_bytree</code>	[0.7, 0.8, 0.9]	
	Random Search	<code>n_estimators</code>	randint (100, 301)	
		<code>learning_rate</code>	uniform(0.01, 0.19)	
		<code>max_depth</code>	randint(3, 8)	
		<code>colsample_bytree</code>	uniform(0.7, 0.2)	
		Bayesian Optimization	<code>n_estimators</code>	Integer(100, 300)
			<code>learning_rate</code>	Real(0.01, 0.2, prior='log-uniform')
Random Forest	Grid Search	<code>n_estimators</code>	[100, 200, 300]	
		<code>max_depth</code>	[None, 10, 20]	
		<code>min_samples_split</code>	[2, 5, 10]	
		<code>min_samples_leaf</code>	[1, 2, 4]	
	Random Search	<code>n_estimators</code>	randint(100, 301)	
		<code>max_depth</code>	Categorical([None, 10, 20])	
		<code>min_samples_split</code>	randint(2, 11)	
		<code>min_samples_leaf</code>	randint(1, 5)	
		Bayesian Optimization	<code>n_estimators</code>	Integer(100, 300)
			<code>max_depth</code>	Categorical([None, 10, 20])
SVM	Grid Search	<code>C</code>	[0.1, 1, 10]	
		<code>kernel</code>	['linear', 'rbf']	
		<code>gamma</code>	['scale', 'auto']	
	Random Search	<code>C</code>	uniform(0.1, 9.9)	
		<code>kernel</code>	Categorical(['linear', 'rbf'])	
		<code>gamma</code>	Categorical(['scale', 'auto'])	
		Bayesian Optimization	<code>C</code>	Real(0.1, 10, prior='log-uniform')

Algoritma	Metode Tuning	Parameter	Rentang Nilai
<b>XGBoost</b>	Grid Search	n_estimators	[100, 200]
		learning_rate	[0.01, 0.1]
		max_depth	[3, 5]
	Random Search	colsample_bytree	randint (100, 201)
		n_estimators	uniform(0.01, 0.09)
		learning_rate	randint(3, 6)
	max_depth	Integer(100, 200)	

<b>Random Forest</b>	Grid Search	n_estimators	randint(100, 301)
		max_depth	Categorical([None, 10])
		min_samples_split	randint(2, 6)
	Random Search	min_samples_leaf	randint(1, 3)
		n_estimators	Integer(100, 200)
		max_depth	Categorical([None, 10, 20])
<b>SVM</b>	Grid Search	min_samples_split	Integer(2, 10)
		min_samples_leaf	[0.1, 1]
	Random Search	n_estimators	['rbf']
		max_depth	uniform(0.1, 0.9)
<b>Bayesian Optimization</b>	Grid Search	min_samples_split	Categorical(['rbf'])
		min_samples_leaf	Real(0.1, 1, prior='log-uniform')
		C	Categorical(['rbf'])
	Random Search	kernel	[100, 200]
		gamma	[0.01, 0.1]
		C	[3, 5]
<b>Bayesian Optimization</b>	Random Search	kernel	randint (100, 201)
		gamma	uniform(0.01, 0.09)
		C	randint(3, 6)
<b>Bayesian Optimization</b>	Random Search	kernel	Integer(100, 200)
		gamma	Real(0.01, 0.1, prior='log-uniform')

b. Ruang Pencarian Hyperparameter pada Dataset Credit Card Fraud

Berbeda dengan dataset Breast Cancer, dataset Credit Card Fraud memiliki ukuran yang jauh lebih besar dan tingkat ketidakseimbangan kelas yang tinggi. Oleh karena itu, ruang pencarian hyperparameter dirancang lebih sempit untuk menjaga efisiensi komputasi dan stabilitas proses pelatihan model.

Pada algoritma XGBoost, hanya beberapa parameter utama yang dieksplorasi, seperti jumlah pohon, learning rate, dan kedalaman pohon. Pembatasan ini dilakukan untuk mengurangi waktu pelatihan sekaligus mencegah overfitting pada data yang sangat tidak seimbang. Pada Random Forest, jumlah pohon dan kedalaman maksimum tetap menjadi fokus utama, dengan pengurangan variasi pada parameter lain. Pendekatan ini dianggap cukup representatif untuk menangkap pola fraud tanpa meningkatkan kompleksitas model secara berlebihan. Sementara itu, pada algoritma SVM, ruang pencarian dibuat paling terbatas dengan hanya menggunakan kernel RBF dan variasi parameter C. Hal ini disebabkan oleh tingginya biaya komputasi SVM pada dataset berskala besar, khususnya ketika menggunakan kernel non-linier.

Tabel 3. Ruang Pencarian Hyperparameter pada Dataset

### 3.5 Penyimpanan Hasil Tuning

Berdasarkan hasil tuning hyperparameter, dataset Breast Cancer yang bersifat berimbang, terlihat bahwa seluruh metode tuning Grid Search, Random Search, dan Bayesian Optimization mampu menghasilkan performa model yang tinggi dan relatif stabil. Model SVM secara konsisten menunjukkan skor terbaik dibandingkan XGBoost dan Random Forest, dengan skor tertinggi diperoleh melalui Bayesian Optimization (0,9799) menggunakan kernel linear dan nilai parameter C yang relatif kecil. Hal ini memperkuat temuan sebelumnya bahwa SVM sangat efektif pada dataset berimbang dengan pola pemisahan kelas yang cukup jelas. Sementara itu, XGBoost dan Random Forest juga menunjukkan performa yang kompetitif dengan variasi parameter yang tidak terlalu ekstrem antar metode tuning, menandakan bahwa eksplorasi ruang hyperparameter yang berbeda tidak menghasilkan perbedaan performa yang signifikan pada dataset ini.

Pada dataset Credit Card Fraud yang bersifat tidak berimbang, hasil tuning menunjukkan variasi parameter dan performa yang lebih beragam antar metode dan model. SVM kembali menunjukkan performa yang sangat tinggi dan stabil pada ketiga metode tuning, dengan skor terbaik mendekati sempurna (0,995) dan penggunaan kernel RBF, yang mengindikasikan kebutuhan model non-linear untuk menangkap pola kompleks pada data fraud. Namun demikian, performa tinggi pada skor terbaik ini perlu diinterpretasikan dengan hati-hati dan dikaitkan dengan metrik evaluasi yang relevan, seperti F1-score dan Precision-Recall AUC, sebagaimana telah dibahas pada analisis sebelumnya.

### 3.6 Evaluasi Model Akhir

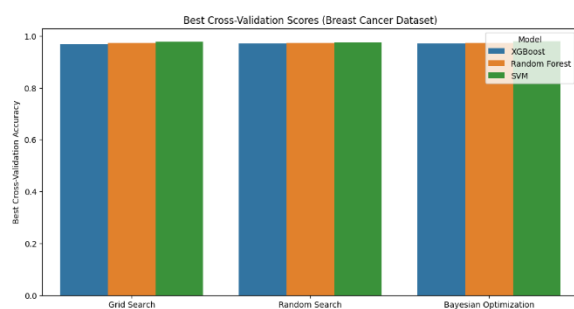
Berdasarkan hasil evaluasi pada dataset Breast Cancer, seluruh model menunjukkan performa yang tinggi dan stabil pada data uji, baik dari sisi accuracy maupun F1-score (macro). Model SVM yang dituning menggunakan Grid Search dan Bayesian Optimization menghasilkan

nilai F1-score macro tertinggi (0,9748), yang sejalan dengan nilai akurasi yang juga tinggi ( $\approx 0,9766$ ). Konsistensi antara akurasi dan F1-score macro ini menunjukkan bahwa model mampu mengklasifikasikan kedua kelas secara seimbang tanpa adanya dominasi performa pada salah satu kelas. XGBoost dan Random Forest juga menunjukkan kinerja yang baik, namun secara konsisten berada sedikit di bawah SVM, baik pada Grid Search, Random Search, maupun Bayesian Optimization. Hal ini mengindikasikan bahwa pada dataset berimbang, pemilihan model dan metode tuning tidak menghasilkan perbedaan performa yang signifikan selama konfigurasi optimal telah tercapai.

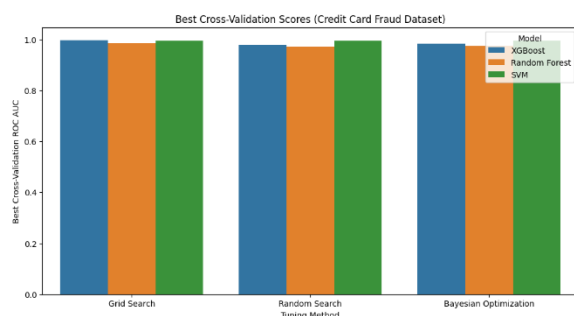
Tabel 4 Hasil evaluasi performa hyperparameter tuning

Dataset Breast Cancer				
Model	Tuning Method	Best Score (CV Accuracy)	Test Set Accuracy	Test Set F1-score (macro)
XGBoost	Grid Search	0.967468354	0.959064327	0.955870968
Random Forest	Grid Search	0.9725	0.935672515	0.931111518
SVM	Grid Search	0.977405063	0.976608187	0.974867725
XGBoost	Random Search	0.969968354	0.959064327	0.95556298
Random Forest	Random Search	0.9725	0.929824561	0.925087617
SVM	Random Search	0.974905063	0.970760234	0.968687054
XGBoost	Bayesian Optimization	0.97	0.941520468	0.937169312
Random Forest	Bayesian Optimization	0.9725	0.935672515	0.931111518
SVM	Bayesian Optimization	0.979905063	0.976608187	0.974867725
Dataset Credit Card Fraud				
Model	Tuning Method	Best Score (CV ROC AUC)	Test Set F1-score (binary)	Test Set Precision-Recall AUC
XGBoost	Grid Search	0.9978	0.102564103	0.75826318
Random Forest	Grid Search	0.9853	0.769784173	0.765010316
SVM	Grid Search	0.9952	0	0.645963411
XGBoost	Random Search	0.9785	0.796992481	0.801726975
Random Forest	Random Search	0.9729	0.795539033	0.776396581
SVM	Random Search	0.9951	0.039735099	0.645953665
XGBoost	Bayesian Optimization	0.9838	0.809338521	0.804944448
Random Forest	Bayesian Optimization	0.9750	0.8	0.781942281
SVM	Bayesian Optimization	0.9951	0	0.646101065

Berdasarkan Tabel 4 yang menyajikan hasil evaluasi pada dataset Breast Cancer, seluruh model menunjukkan performa yang tinggi dan relatif stabil pada data uji. Model SVM yang dituning menggunakan *Grid Search* dan *Bayesian Optimization* menghasilkan nilai Test Set F1-score (macro) tertinggi (0,9748), sejalan dengan akurasi yang juga tinggi ( $\approx 0,9766$ ). Hal ini menunjukkan bahwa SVM mampu mengklasifikasikan kedua kelas secara seimbang dan konsisten. *XGBoost* dan *Random Forest* juga menunjukkan performa yang baik, namun secara umum berada sedikit di bawah SVM. Konsistensi nilai accuracy dan F1-score macro pada seluruh metode tuning mengindikasikan bahwa pada dataset berimbang, perbedaan metode hyperparameter tuning tidak memberikan dampak signifikan terhadap kinerja akhir model terbaik.



Gambar 1. Cross Validation dataset Breast Cancer



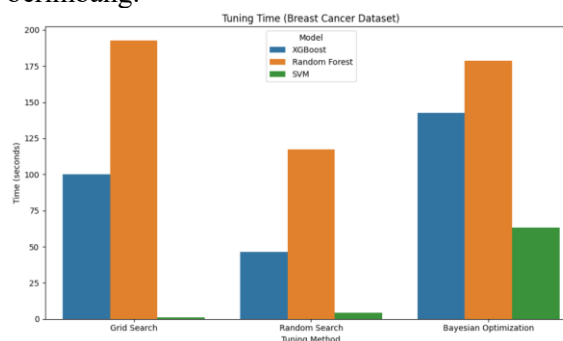
Gambar 2 Cross Validation dataset Credit Card Fraud

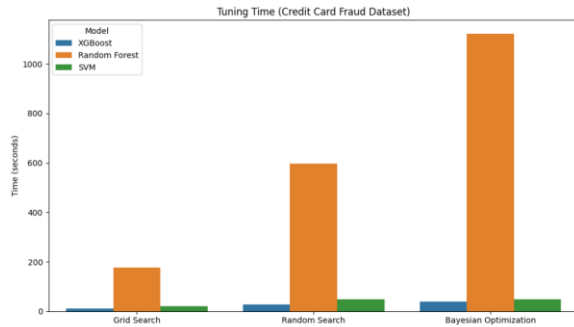
Penggunaan F1-score (macro) sebagai metrik evaluasi utama pada dataset Breast Cancer dipilih karena metrik ini memberikan bobot yang sama pada setiap kelas, sehingga mampu

merepresentasikan performa model secara adil pada kondisi kelas yang relatif seimbang. Dibandingkan hanya menggunakan akurasi, F1-score macro lebih informatif karena mempertimbangkan keseimbangan antara precision dan recall pada masing-masing kelas, sehingga sesuai untuk mengevaluasi kualitas klasifikasi secara menyeluruh.

Sebaliknya, pada dataset Credit Card Fraud yang bersifat sangat tidak berimbang, hasil pada menunjukkan perbedaan yang cukup jelas antara skor cross-validation dan performa pada data uji. Beberapa model, khususnya SVM, memperoleh nilai CV ROC AUC yang sangat tinggi, namun menghasilkan Test Set F1-score (binary) yang sangat rendah bahkan nol, yang menandakan kegagalan model dalam mendeteksi kelas minoritas. Model XGBoost dan Random Forest yang dituning menggunakan Random Search dan Bayesian Optimization justru menunjukkan performa yang lebih baik pada Test Set Precision-Recall AUC, dengan nilai tertinggi dicapai oleh XGBoost dan Bayesian Optimization (0,8049).

Pemilihan Precision-Recall AUC sebagai metrik utama pada dataset Credit Card Fraud didasarkan pada kemampuannya dalam mengevaluasi performa model terhadap kelas minoritas secara lebih representatif dibandingkan akurasi atau ROC AUC. Precision-Recall AUC menekankan keseimbangan antara precision dan recall pada kelas fraud, sehingga lebih sesuai untuk skenario deteksi penipuan dengan distribusi kelas yang timpang. Oleh karena itu, hasil evaluasi pada tabel menunjukkan bahwa metode tuning dan model yang berbeda dapat memberikan dampak yang lebih signifikan pada dataset tidak berimbang dibandingkan dataset berimbang.





Gambar 3 Perbandingan waktu tuning pada kedua dataset

## 4. Kesimpulan dan Saran

### 4.1 Kesimpulan

Bahwa performa model sangat bergantung pada karakteristik dataset. SVM cenderung unggul pada dataset dengan dimensi fitur yang relatif rendah dan seimbang seperti Breast Cancer, sementara model berbasis tree (XGBoost, Random Forest) menunjukkan performa yang kuat pada dataset dengan dimensi dan kompleksitas yang lebih tinggi serta ketidakseimbangan kelas yang parah seperti Credit Card Fraud ketika di-tune dengan metode yang tepat.

Untuk dataset yang tidak seimbang, seperti Credit Card Fraud, metrik evaluasi yang berfokus pada kelas minoritas (seperti F1-score binary dan Precision-Recall AUC) jauh lebih informatif dibandingkan akurasi. Hasil menunjukkan bahwa tuning hyperparameter sangat penting untuk meningkatkan kemampuan model dalam mendeteksi kelas minoritas ini.

Analisis ini memberikan gambaran komparatif mengenai efektivitas berbagai kombinasi model dan metode tuning, termasuk pertimbangan efisiensi komputasi. Untuk aplikasi praktis, pemilihan model dan metode tuning terbaik harus mempertimbangkan tidak hanya metrik performa pada set pengujian, tetapi juga aspek lain seperti kebutuhan komputasi, waktu yang tersedia untuk tuning, dan interpretasi model. Bayesian Optimization, khususnya, menunjukkan potensi untuk menjadi pilihan yang menarik ketika menyeimbangkan kebutuhan akan performa optimal dan efisiensi waktu tuning.

### 4.2 Saran

Pemilihan metrik evaluasi sangat penting dan harus selaras dengan karakteristik dataset, khususnya ketidakseimbangan kelas, untuk secara akurat menilai kinerja model (misalnya, menggunakan ROC AUC/Precision-Recall AUC untuk dataset yang tidak seimbang seperti Credit Card Fraud).

Terdapat pertukaran antara waktu penyetalan dan kinerja model, terutama pada dataset yang lebih besar dan kompleks. Optimasi Bayesian tampak menjanjikan untuk menyeimbangkan kinerja dan efisiensi komputasi saat menyetal model kompleks pada dataset tersebut.

## DAFTAR PUSTAKA

- Bischl, Bernd, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, et al. 2023. "Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 13 (2): e1484.
- Bishop, Christopher M, and Hugh Bishop. 2023. *Deep Learning: Foundations and Concepts*. Springer Nature.
- Dasgupta, Subhasis, and Jaydip Sen. 2024. "A Comparative Study of Hyperparameter Tuning Methods." <https://doi.org/10.48550/arXiv.2408.16425>.
- Johnson, Justin, and Taghi Khoshgoftaar. 2019. "Survey on Deep Learning with Class Imbalance." *Journal of Big Data* 6: 27. <https://doi.org/10.1186/s40537-019-0192-5>.
- Kaur, Harsurinder, Husanbir Pannu, and Avleen Malhi. 2019. "A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions." *ACM Computing Surveys* 52: 1–36. <https://doi.org/10.1145/3343440>.
- Michalakopoulos, Vasilis, Ioannis Papias, Efstathios Sarantinopoulos, Elissaios

- Sarmas, Vangelis Marinakis, and Dimitris Askounis. 2025. "A Hyperparameter-Space Clustering Methodology of Residential Electricity Loads." *Applied Soft Computing* 181: 113497. <https://doi.org/10.1016/j.asoc.2025.113497>
- Morales-Hernández, Alejandro, Inneke Nieuwenhuysse, and Sebastian Gonzalez. 2022. "A Survey on Multi-Objective Hyperparameter Optimization Algorithms for Machine Learning." *Artificial Intelligence Review* 56: 1–51. <https://doi.org/10.1007/s10462-022-10359-2>
- Parekh, Sagar, Heramb Nemlekar, and Dylan Losey. 2025. "Towards Balanced Behavior Cloning from Imbalanced Datasets." <https://doi.org/10.48550/arXiv.2508.06319>
- Tiastama, Sheren Afryan, and Indra Budi. 2024. "Perbandingan Random Search Dan Algoritma Genetika Dalam Penyetelan Hyperparameter XGBoost Pada Retail Sales Forecasting." *Indonesian Journal of Computer Science* 13 (4): 6602–13. <https://doi.org/https://doi.org/10.33022/ijcs.v13i4.4285>
- Yang, Li, and Abdallah Shami. 2020. "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice." *Neurocomputing* 415: 295–316. <https://doi.org/10.1016/j.neucom.2020.07.061>