

OPTIMISASI HIPERPARAMETER ALGORITMA MACHINE LEARNING MENGGUNAKAN GRID SEARCH UNTUK KLASIFIKASI KELANGSUNGAN HIDUP PENUMPANG KAPAL

(Hyperparameter Optimisation of Machine Learning Algorithms Using Grid Search for Ship Passenger Survival Classification)

Fadli Husein Wattiheluw¹, Nizmah Ahmad Bahasoan²

^{1,2} Institut Teknologi dan Bisnis Stikom, Jl. A. Y. Patty No. 98, Kota Ambon, Maluku 97126, Indonesia

*Email korespondensi: fadliwattiheluw1994@gmail.com

Abstract

Passenger survival prediction provides a relevant classification setting for evaluating the ability of machine learning algorithms to identify risk patterns from passenger characteristics. This study aims to evaluate the effects of hyperparameter optimisation on classification performance and to compare the effectiveness of Grid Search, Random Search, and Bayesian Optimisation in identifying optimal model configurations. The Titanic dataset was employed as a case study, with the analytical workflow comprising data preprocessing, missing-value treatment, categorical variable encoding, normalisation, and class balancing using the Synthetic Minority Over-sampling Technique (SMOTE). Three machine learning algorithms—K-Nearest Neighbours (KNN), Logistic Regression, and Support Vector Machine (SVM)—were evaluated using accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The results demonstrate that hyperparameter optimisation improves model performance and consistency, with KNN and SVM exhibiting more pronounced improvements, whereas Logistic Regression shows comparatively limited changes in performance. Comparison across optimisation approaches indicates that Grid Search can produce competitive hyperparameter configurations through systematic exploration of a predefined parameter space. These findings demonstrate that algorithm selection and hyperparameter optimisation strategy are important components in developing accurate and reliable classification models. Methodologically, the proposed approach provides a potential foundation for developing data-driven predictive models to support risk analysis and decision-support applications in the maritime transport sector.

Keywords: classification; Grid Search; hyperparameters; machine learning; optimisation; passenger survival.

Abstrak

Prediksi kelangsungan hidup penumpang merupakan salah satu aplikasi klasifikasi yang relevan untuk mengevaluasi kemampuan algoritma machine learning dalam mengidentifikasi pola risiko berbasis karakteristik penumpang. Penelitian ini bertujuan mengevaluasi pengaruh optimisasi hiperparameter terhadap kinerja model klasifikasi serta membandingkan efektivitas Grid Search,

Random Search, dan Bayesian Optimisation dalam memperoleh konfigurasi model yang optimal. Dataset Titanic digunakan sebagai studi kasus dengan tahapan analisis meliputi prapemrosesan data, penanganan nilai hilang, pengkodean variabel kategorikal, normalisasi, dan penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE). Tiga algoritma, yaitu K-Nearest Neighbours (KNN), Regresi Logistik, dan Support Vector Machine (SVM), dievaluasi menggunakan akurasi, presisi, recall, F1-score, dan Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Hasil penelitian menunjukkan bahwa optimisasi hiperparameter meningkatkan kinerja dan konsistensi model, terutama pada KNN dan SVM, sedangkan Regresi Logistik menunjukkan perubahan kinerja yang relatif terbatas. Perbandingan antarmetode optimisasi menunjukkan bahwa Grid Search mampu menghasilkan konfigurasi hiperparameter yang kompetitif melalui eksplorasi sistematis terhadap ruang parameter yang telah ditentukan. Temuan ini menunjukkan bahwa pemilihan algoritma dan strategi optimisasi hiperparameter merupakan komponen penting dalam membangun model klasifikasi yang akurat dan andal. Secara metodologis, pendekatan ini berpotensi mendukung pengembangan model prediktif berbasis data untuk analisis risiko dan sistem pendukung keputusan pada sektor transportasi maritim.

Kata kunci: klasifikasi; Grid Search; hiperparameter; machine learning; optimisasi; kelangsungan hidup penumpang

I. Pendahuluan

Perkembangan kecerdasan buatan, khususnya pembelajaran mesin (machine learning), telah memperluas kemampuan analisis data dalam mengidentifikasi pola, membangun model klasifikasi, dan menghasilkan prediksi berdasarkan data historis [1]. Meskipun demikian, kinerja model pembelajaran mesin tidak hanya ditentukan oleh karakteristik data dan pemilihan algoritma, tetapi juga oleh konfigurasi hiperparameter yang mengendalikan proses pembelajaran. Konfigurasi yang kurang tepat dapat menghasilkan model dengan kemampuan generalisasi yang rendah, sedangkan optimisasi hiperparameter yang dilakukan secara sistematis berpotensi meningkatkan kinerja prediktif dan konsistensi model pada data yang belum pernah diamati.

Dalam konteks transportasi maritim, keselamatan penumpang merupakan isu penting karena kecelakaan kapal dapat menimbulkan konsekuensi serius berupa korban jiwa, kerugian ekonomi, dan gangguan terhadap sistem transportasi [2], [3]. Pemanfaatan analisis prediktif berbasis pembelajaran mesin menawarkan pendekatan berbasis data untuk mengidentifikasi pola yang berkaitan dengan risiko dan kelangsungan hidup penumpang [4]. Namun, kemampuan suatu model untuk menghasilkan klasifikasi yang andal sangat bergantung pada kualitas data, representasi fitur, pemilihan algoritma, strategi validasi, dan konfigurasi hiperparameter. Oleh sebab itu, evaluasi model tidak cukup dilakukan hanya berdasarkan pemilihan algoritma, tetapi perlu mempertimbangkan proses optimisasi yang terstruktur.

Dataset Titanic menyediakan studi kasus klasifikasi yang memungkinkan evaluasi hubungan antara karakteristik penumpang dan status kelangsungan hidup. Atribut seperti usia, jenis kelamin, kelas penumpang, tarif, dan sejumlah karakteristik perjalanan dapat digunakan sebagai prediktor untuk membedakan penumpang yang selamat dan tidak selamat. Dalam penelitian ini, tiga algoritma pembelajaran terawasi, yaitu K-Nearest

Neighbours (KNN), Regresi Logistik (Logistic Regression/LR), dan Support Vector Machine (SVM), digunakan karena merepresentasikan karakteristik pembelajaran yang berbeda. KNN mengandalkan kedekatan antarsampel, Regresi Logistik membangun fungsi keputusan probabilistik, sedangkan SVM menentukan batas keputusan berdasarkan pemisahan kelas dalam ruang fitur. Perbedaan karakteristik tersebut memungkinkan analisis mengenai sejauh mana masing-masing algoritma merespons optimisasi hiperparameter.

Salah satu pendekatan yang digunakan untuk optimisasi hiperparameter adalah Grid Search, yaitu pencarian sistematis terhadap kombinasi nilai hiperparameter yang telah ditentukan sebelumnya [5], [6]. Melalui evaluasi setiap kombinasi berdasarkan prosedur validasi dan metrik kinerja yang konsisten, Grid Search memungkinkan identifikasi konfigurasi yang memberikan kinerja terbaik dalam ruang pencarian yang ditetapkan. Meskipun demikian, efektivitas pendekatan ini bergantung pada desain ruang hiperparameter, strategi validasi, kompleksitas algoritma, dan karakteristik dataset [7]. Oleh karena itu, keunggulan Grid Search tidak dapat diasumsikan secara apriori, tetapi perlu diuji secara empiris dan, apabila relevan, dibandingkan dengan strategi pencarian alternatif seperti Random Search dan Bayesian Optimisation.

Tantangan lain dalam klasifikasi kelangsungan hidup penumpang berkaitan dengan kualitas dan karakteristik data. Nilai hilang, variabel kategorikal, perbedaan skala antarfitur, fitur yang kurang informatif, dan kemungkinan ketidakseimbangan kelas dapat memengaruhi proses pembelajaran dan hasil evaluasi [8]. Penanganan aspek-aspek tersebut memerlukan alur prapemrosesan yang konsisten. Teknik Synthetic Minority Over-sampling Technique (SMOTE) dapat digunakan pada data pelatihan apabila ditemukan ketidakseimbangan kelas [9], sedangkan Recursive Feature Elimination (RFE) dapat digunakan untuk mengevaluasi dan mempertahankan fitur yang relevan terhadap proses klasifikasi [10]. Seluruh transformasi berbasis data perlu ditempatkan dalam prosedur pelatihan dan validasi yang tepat agar tidak terjadi kebocoran informasi (data leakage) dari data pengujian ke dalam proses pembentukan model.

Meskipun klasifikasi kelangsungan hidup penumpang Titanic telah banyak diteliti menggunakan berbagai algoritma pembelajaran mesin, masih terdapat ruang untuk mengevaluasi secara lebih terstruktur pengaruh optimisasi hiperparameter terhadap algoritma dengan karakteristik pembelajaran yang berbeda. Sejumlah penelitian terdahulu lebih menekankan perbandingan akurasi antarmodel atau interpretasi faktor yang berkaitan dengan kelangsungan hidup, sementara evaluasi mengenai perubahan kinerja model sebelum dan setelah optimisasi hiperparameter belum selalu menjadi fokus utama. Selain itu, penggunaan prosedur prapemrosesan, ruang hiperparameter, strategi pencarian, dan metrik evaluasi yang berbeda antarpublikasi menyulitkan penilaian mengenai sejauh mana peningkatan kinerja benar-benar berasal dari optimisasi hiperparameter.

Berdasarkan kesenjangan tersebut, penelitian ini mengembangkan kerangka evaluasi komparatif untuk menganalisis pengaruh optimisasi hiperparameter terhadap kinerja KNN, Regresi Logistik, dan SVM dalam klasifikasi kelangsungan hidup penumpang Titanic. Grid Search digunakan sebagai pendekatan utama untuk mengidentifikasi

konfigurasi hiperparameter optimal, dengan evaluasi berdasarkan akurasi, presisi, recall, F1-score, dan Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Apabila Random Search dan Bayesian Optimisation juga diterapkan dalam eksperimen, keduanya digunakan sebagai pembanding untuk menilai efektivitas relatif strategi pencarian hiperparameter berdasarkan kerangka evaluasi yang sama.

Kontribusi penelitian ini terletak pada tiga aspek. Pertama, penelitian mengevaluasi sensitivitas tiga algoritma dengan karakteristik pembelajaran berbeda terhadap optimisasi hiperparameter. Kedua, penelitian menerapkan kerangka prapemrosesan, optimisasi, dan evaluasi yang konsisten sehingga perubahan kinerja antarmodel dapat dibandingkan secara lebih objektif. Ketiga, penelitian mengidentifikasi kombinasi algoritma dan konfigurasi hiperparameter yang memberikan kinerja klasifikasi paling andal pada dataset yang dianalisis. Dengan demikian, kontribusi penelitian tidak semata-mata terletak pada penggunaan Grid Search, tetapi pada evaluasi empiris mengenai bagaimana strategi optimisasi memengaruhi kinerja dan konsistensi model klasifikasi.

Penelitian ini bertujuan untuk: (1) mengevaluasi kinerja KNN, Regresi Logistik, dan SVM dalam mengklasifikasikan kelangsungan hidup penumpang; (2) mengidentifikasi konfigurasi hiperparameter terbaik melalui Grid Search; (3) mengukur perubahan kinerja model sebelum dan setelah optimisasi berdasarkan beberapa metrik evaluasi; serta (4) menentukan model yang memberikan keseimbangan kinerja paling baik untuk dataset yang dianalisis. Hasil penelitian diharapkan memberikan bukti empiris mengenai pentingnya optimisasi hiperparameter dalam pengembangan model klasifikasi yang akurat dan konsisten, sekaligus memberikan kerangka metodologis yang dapat direplikasi pada permasalahan klasifikasi berbasis data lainnya.

II. Penelitian Terkait

Penelitian mengenai klasifikasi kelangsungan hidup penumpang menggunakan pembelajaran mesin telah berkembang melalui penerapan berbagai algoritma pembelajaran terawasi. Pendekatan yang umum digunakan meliputi Regresi Logistik, Decision Tree, K-Nearest Neighbours (KNN), Support Vector Machine (SVM), dan berbagai metode ensemble. Secara umum, penelitian terdahulu berfokus pada dua orientasi utama, yaitu membandingkan kinerja prediktif antarmodel dan mengidentifikasi atribut penumpang yang berkontribusi terhadap hasil klasifikasi. Kinerja model umumnya dievaluasi menggunakan akurasi, presisi, recall, F1-score, dan ukuran diskriminasi lainnya. Namun, perhatian terhadap optimisasi hiperparameter dan konsistensi prosedur validasi masih bervariasi antarpublikasi.

Shaik dkk. [11] menerapkan Regresi Logistik, Decision Tree, dan KNN untuk mengembangkan model klasifikasi kelangsungan hidup penumpang. Studi tersebut menunjukkan bahwa algoritma dengan mekanisme pembelajaran berbeda dapat menghasilkan performa yang berbeda pada permasalahan klasifikasi penumpang. Meskipun memberikan dasar penting bagi perbandingan algoritma, aspek optimisasi hiperparameter belum menjadi fokus utama penelitian tersebut. Dengan demikian, masih diperlukan evaluasi mengenai apakah konfigurasi hiperparameter yang diperoleh melalui prosedur

pencarian sistematis mampu menghasilkan peningkatan kinerja dibandingkan konfigurasi awal atau baseline.

Bisht dkk. [12] membandingkan sejumlah algoritma dalam klasifikasi kelangsungan hidup penumpang Titanic dan melaporkan bahwa CatBoost mencapai akurasi sebesar 78,41%. Model tersebut menggunakan sejumlah karakteristik penumpang, termasuk usia, kelas perjalanan, jenis kelamin, pelabuhan keberangkatan, dan tarif. Hasil penelitian menunjukkan potensi algoritma berbasis ensemble dalam menangkap hubungan kompleks antarfaktor prediktor. Namun, apabila studi tersebut tidak menerapkan prosedur optimisasi hiperparameter secara eksplisit, hasil perbandingan tersebut belum sepenuhnya menjelaskan sejauh mana konfigurasi hiperparameter memengaruhi performa relatif masing-masing algoritma. Oleh karena itu, evaluasi terkontrol antara konfigurasi awal dan hasil optimisasi tetap relevan untuk dilakukan.

Dasgupta dkk. [13] mengevaluasi hubungan sejumlah karakteristik, termasuk usia, jenis kelamin, dan kelas sosial, dengan probabilitas kelangsungan hidup penumpang. Studi tersebut memperlihatkan bahwa karakteristik demografis dan sosial memiliki kontribusi penting terhadap pola klasifikasi. Temuan ini memperkuat relevansi analisis berbasis fitur dalam memahami struktur data Titanic. Namun, orientasi penelitian yang lebih menekankan interpretasi atribut menyebabkan aspek teknis terkait sensitivitas algoritma terhadap konfigurasi hiperparameter belum menjadi fokus utama.

Sintesis terhadap penelitian terdahulu menunjukkan bahwa permasalahan utama bukan lagi sekadar apakah pembelajaran mesin dapat digunakan untuk mengklasifikasi kelangsungan hidup penumpang, tetapi bagaimana kinerja model berubah ketika algoritma dievaluasi menggunakan konfigurasi hiperparameter yang dioptimalkan dalam kerangka validasi yang konsisten. Penggunaan konfigurasi default atau penyesuaian manual dapat menghasilkan estimasi kinerja yang belum optimal dan menyulitkan perbandingan yang adil antaralgoritma. Selain itu, algoritma yang berbeda tidak memiliki tingkat sensitivitas yang sama terhadap hiperparameter sehingga manfaat optimisasi perlu dibuktikan secara empiris untuk setiap model.

Berangkat dari kesenjangan tersebut, penelitian ini menerapkan Grid Search untuk mengoptimalkan hiperparameter KNN, Regresi Logistik, dan SVM dalam satu kerangka eksperimen yang konsisten. Ketiga algoritma dipilih karena merepresentasikan mekanisme klasifikasi yang berbeda sehingga memungkinkan evaluasi terhadap variasi respons model terhadap proses optimisasi. Kinerja model sebelum dan setelah optimisasi dibandingkan menggunakan akurasi, presisi, recall, F1-score, dan AUC-ROC. Apabila Random Search dan Bayesian Optimisation turut digunakan, kedua pendekatan tersebut diposisikan sebagai pembanding untuk menentukan apakah pencarian sistematis melalui Grid Search memberikan keuntungan empiris dibandingkan strategi optimisasi alternatif.

III. Metode Penelitian

3.1 Pengumpulan dan Persiapan Dataset

Dataset yang digunakan adalah Titanic, yang memuat karakteristik penumpang RMS *Titanic* dan status kelangsungan hidup setelah peristiwa tenggelamnya kapal pada tahun 1912. Dataset terdiri atas 891 observasi dan 12 atribut, dengan variabel *Survived* sebagai target klasifikasi biner yang merepresentasikan status penumpang selamat atau tidak selamat. Sejumlah atribut penjelas mencakup karakteristik demografis dan perjalanan penumpang, antara lain usia (*Age*), jenis kelamin (*Sex*), kelas penumpang (*Pclass*), tarif (*Fare*), jumlah anggota keluarga tertentu yang bepergian bersama penumpang, serta pelabuhan keberangkatan (*Embarked*).

Dataset Titanic dipilih sebagai *benchmark case study* karena menyediakan permasalahan klasifikasi biner dengan kombinasi fitur numerik dan kategorikal, nilai hilang, serta karakteristik data yang memungkinkan pengujian beberapa tahapan prapemrosesan dan optimisasi model. Oleh karena itu, dataset ini digunakan terutama untuk mengevaluasi efektivitas prosedur klasifikasi dan optimisasi hiperparameter, bukan sebagai representasi langsung terhadap seluruh kompleksitas sistem keselamatan transportasi maritim kontemporer. Interpretasi dan generalisasi hasil penelitian selanjutnya dibatasi pada karakteristik dataset yang dianalisis.

Prapemrosesan data dilakukan untuk meningkatkan konsistensi dan kesiapan data sebelum proses pemodelan. Tahapan ini mencakup pemeriksaan duplikasi dan nilai hilang, penanganan *missing values*, pengkodean variabel kategorikal, serta penskalaan fitur numerik apabila diperlukan oleh karakteristik algoritma. Variabel kategorikal ditransformasikan ke dalam representasi numerik yang sesuai, sedangkan normalisasi atau standarisasi diterapkan terutama untuk algoritma yang sensitif terhadap skala fitur, seperti KNN dan SVM.

Data selanjutnya dipisahkan menjadi data pelatihan (*training set*) dan data pengujian (*test set*). Data pengujian dipertahankan sebagai data independen dan tidak digunakan selama proses pemilihan hiperparameter. Apabila distribusi kelas pada variabel target menunjukkan ketidakseimbangan yang memerlukan penanganan, *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan hanya pada data pelatihan dan, dalam proses validasi silang, hanya pada bagian pelatihan dari setiap *fold*. Strategi ini diperlukan untuk mencegah terbentuknya sampel sintetis berdasarkan informasi dari data validasi atau data uji sehingga risiko *data leakage* dapat diminimalkan.

3.2 Implementasi Algoritma Pembelajaran Mesin

Tiga algoritma klasifikasi digunakan dalam penelitian ini, yaitu Regresi Logistik (LR), SVM, dan KNN. Pemilihan ketiga algoritma didasarkan pada perbedaan prinsip pembelajarannya sehingga memungkinkan evaluasi pengaruh optimisasi hiperparameter pada model dengan karakteristik yang berbeda. Regresi Logistik digunakan sebagai model klasifikasi linear berbasis probabilitas; SVM membentuk *decision boundary* dengan memaksimalkan margin pemisah antarkelas dan dapat memanfaatkan fungsi *kernel* untuk

menangani hubungan nonlinier; sedangkan KNN melakukan klasifikasi berdasarkan kedekatan sampel terhadap sejumlah tetangga terdekat dalam ruang fitur.

Pada tahap awal, setiap algoritma dibangun menggunakan konfigurasi *baseline* yang konsisten dengan implementasi perangkat lunak yang digunakan. Model *baseline* berfungsi sebagai titik acuan untuk mengukur perubahan kinerja setelah optimisasi hiperparameter. Seluruh model dilatih menggunakan data pelatihan yang telah melalui prosedur prapemrosesan yang sesuai, sedangkan data uji tidak digunakan dalam proses pembentukan maupun pemilihan konfigurasi model.

Kinerja awal masing-masing model kemudian dihitung menggunakan metrik yang sama dengan evaluasi model hasil optimisasi. Perbandingan tersebut digunakan untuk menentukan apakah optimisasi hiperparameter menghasilkan peningkatan kinerja yang substantif dibandingkan model *baseline*. Dengan desain ini, pengaruh optimisasi dapat dievaluasi secara lebih transparan untuk setiap algoritma.

2.3 Optimisasi Hiperparameter Menggunakan Grid Search

Optimisasi hiperparameter dilakukan menggunakan *Grid Search*, yaitu prosedur pencarian sistematis yang mengevaluasi kombinasi nilai hiperparameter dari ruang pencarian (*search space*) yang telah ditentukan. Untuk setiap algoritma, ruang hiperparameter disusun berdasarkan parameter yang secara teoritis berpengaruh terhadap perilaku model. Pada KNN, parameter yang dapat dioptimalkan antara lain jumlah tetangga (*n_neighbors*), skema pembobotan (*weights*), dan metrik jarak. Pada SVM, ruang pencarian dapat mencakup parameter regularisasi (*C*), fungsi *kernel*, dan parameter *gamma*. Sementara itu, pada Regresi Logistik, parameter yang dievaluasi dapat mencakup kekuatan regularisasi (*C*), jenis *penalty*, dan *solver* yang kompatibel. Setiap kombinasi hiperparameter dievaluasi melalui prosedur validasi silang (*cross-validation*) pada data pelatihan. Model dengan skor validasi terbaik berdasarkan metrik utama yang telah ditetapkan dipilih sebagai konfigurasi optimal. Setelah konfigurasi terbaik diperoleh, model dilatih kembali menggunakan data pelatihan dan selanjutnya dievaluasi satu kali pada data uji independen. Pendekatan ini menjaga pemisahan antara proses pemilihan model dan evaluasi akhir sehingga estimasi performa generalisasi tidak bergantung pada data yang digunakan selama proses optimisasi.

Grid Search dipilih karena menyediakan prosedur pencarian yang transparan dan dapat direplikasi pada ruang hiperparameter diskret yang telah ditentukan. Namun, metode ini tidak diasumsikan secara otomatis menghasilkan optimum global di luar ruang pencarian yang didefinisikan. Kualitas konfigurasi yang diperoleh bergantung pada cakupan ruang hiperparameter, strategi validasi, metrik optimisasi, dan sumber daya komputasi yang tersedia. Oleh karena itu, hasil penelitian diinterpretasikan sebagai konfigurasi terbaik **dalam ruang pencarian yang diuji**, bukan sebagai konfigurasi optimal secara universal.

2.4. Evaluasi Model dan Analisis Hasil

Evaluasi akhir dilakukan terhadap model *baseline* dan model hasil optimisasi menggunakan data uji yang tidak dilibatkan dalam proses pelatihan maupun pencarian

hiperparameter. Kinerja model diukur menggunakan akurasi, presisi, *recall*, F1-score, dan *Area Under the Receiver Operating Characteristic Curve* (AUC-ROC). Penggunaan beberapa metrik dilakukan agar penilaian model tidak hanya bergantung pada proporsi prediksi yang benar, tetapi juga mempertimbangkan kemampuan model dalam mengidentifikasi kelas positif dan keseimbangan antara presisi dan sensitivitas.

Akurasi digunakan untuk mengukur proporsi keseluruhan prediksi yang diklasifikasikan dengan benar. Presisi menunjukkan proporsi prediksi positif yang benar, sedangkan *recall* mengukur kemampuan model dalam mengidentifikasi seluruh observasi yang benar-benar termasuk dalam kelas positif. F1-score dihitung sebagai rata-rata harmonik antara presisi dan *recall*, sehingga memberikan ukuran kinerja yang mempertimbangkan keseimbangan keduanya. AUC-ROC digunakan untuk mengevaluasi kemampuan diskriminatif model pada berbagai ambang klasifikasi.

Analisis komparatif dilakukan pada dua tingkat. Pertama, kinerja setiap algoritma dibandingkan antara kondisi *baseline* dan setelah optimisasi hiperparameter untuk mengukur besarnya perubahan performa. Kedua, model hasil optimisasi dibandingkan antaralgoritma untuk mengidentifikasi konfigurasi yang memberikan keseimbangan kinerja terbaik berdasarkan keseluruhan metrik evaluasi. Model terbaik tidak ditentukan hanya berdasarkan satu nilai akurasi, tetapi dengan mempertimbangkan konsistensi akurasi, presisi, *recall*, F1-score, dan AUC-ROC sesuai dengan tujuan klasifikasi.

IV. Hasil dan Pembahasan

4.1 Karakteristik Dataset dan Hasil Prapemrosesan Data

Dataset Titanic yang digunakan dalam penelitian ini terdiri atas 891 observasi dan 12 variabel yang merepresentasikan karakteristik demografis, sosial, dan perjalanan penumpang, dengan *Survived* sebagai variabel target biner. Dataset tersebut diposisikan sebagai *benchmark case study* untuk mengevaluasi efektivitas algoritma klasifikasi dan sensitivitasnya terhadap optimisasi hiperparameter, bukan sebagai representasi langsung dari keseluruhan kompleksitas sistem keselamatan transportasi maritim kontemporer. Pembatasan interpretasi ini penting karena karakteristik historis, struktur variabel, dan ukuran dataset Titanic berbeda dari data operasional keselamatan kapal modern. Karakteristik variabel penelitian secara lengkap disajikan pada **Tabel 1**.

Tabel 1. Ruang Pencarian dan Fungsi Hiperparameter pada Algoritma Pembelajaran Mesin

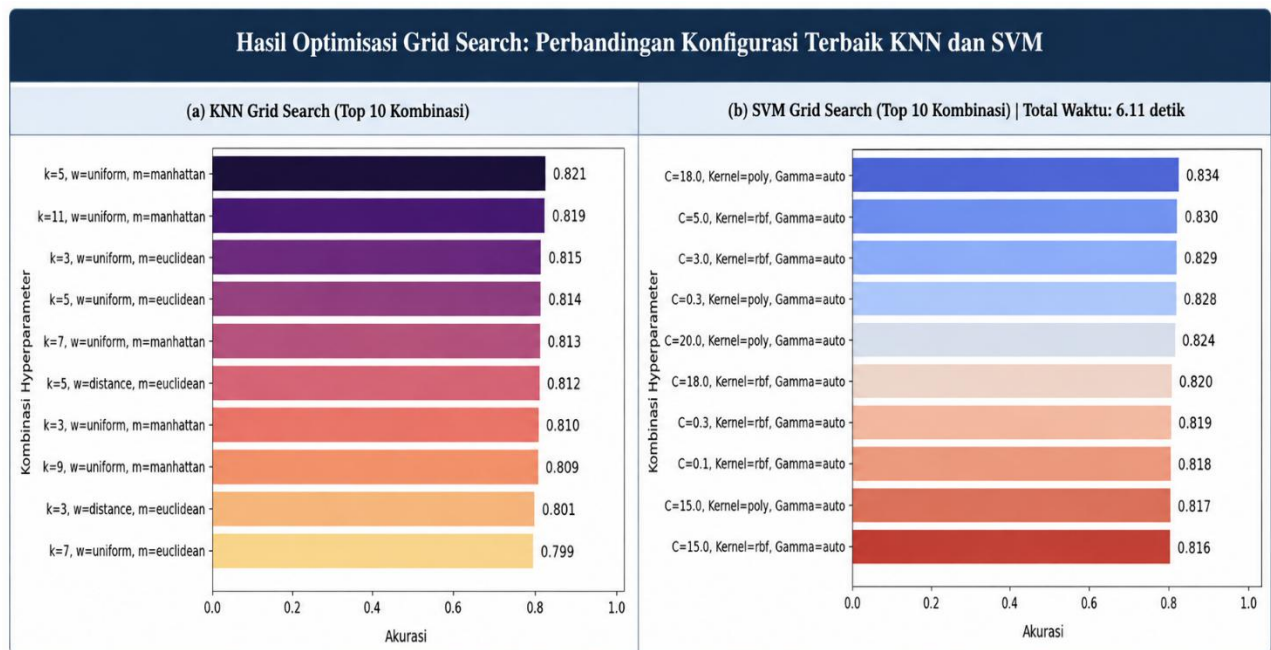
Algoritma	Hiperparameter	Fungsi dalam Model	Nilai/Ruang Pencarian
KNN	n_neighbors	Menentukan jumlah tetangga terdekat	3, 5, 7, 9, 11
	weights	Menentukan pembobotan kontribusi tetangga	uniform, distance

Algoritma	Hiperparameter	Fungsi dalam Model	Nilai/Ruang Pencarian
	metric	Menentukan ukuran jarak antarobservasi	euclidean, manhattan
Regresi Logistik	C	Mengendalikan kekuatan regularisasi	0,1; 0,3; 0,7; 1; 3; 5; 7; 10; 12; 15; 18; 20
	penalty	Menentukan jenis regularisasi	l1, l2
	solver	Menentukan algoritma optimisasi numerik	liblinear, saga
SVM	C	Mengendalikan kompromi margin dan penalti kesalahan	0,1; 0,3; 0,7; 1; 3; 5; 7; 10; 12; 15; 18; 20
	kernel	Menentukan fungsi transformasi ruang fitur	rbf, linear, poly
	gamma*	Mengendalikan pengaruh sampel pada <i>kernel</i> yang relevan	auto

Catatan: gamma='auto' merupakan konfigurasi tetap apabila hanya nilai tersebut yang digunakan; oleh karena itu, parameter ini tidak diperlakukan sebagai dimensi pencarian yang divariasikan.

Eksplorasi awal mengidentifikasi nilai hilang terutama pada variabel *Age* dan *Embarked*. Nilai hilang pada *Age* ditangani menggunakan median, sedangkan *Embarked* diimputasi menggunakan modus. Setelah proses imputasi, kedua variabel tidak lagi mengandung nilai hilang. Selanjutnya, variabel kategorikal ditransformasikan menjadi representasi numerik, dengan *Sex* dikodekan secara biner dan *Embarked* ditransformasikan melalui *one-hot encoding*. Tahapan tersebut memastikan bahwa seluruh prediktor dapat diproses secara konsisten oleh K-Nearest Neighbours (KNN), Regresi Logistik (LR), dan Support Vector Machine (SVM).

Fitur numerik selanjutnya ditransformasikan menggunakan **Min-Max normalisation** pada rentang 0–1. Normalisasi merupakan tahapan penting terutama bagi KNN dan SVM karena kedua algoritma sensitif terhadap perbedaan skala fitur. Pada KNN, skala memengaruhi perhitungan jarak antarobservasi, sedangkan pada SVM skala fitur dapat memengaruhi pembentukan *decision boundary*. Dengan demikian, rangkaian imputasi, pengkodean, dan normalisasi bukan sekadar prosedur teknis, tetapi merupakan mekanisme untuk membangun ruang fitur yang lebih konsisten sebelum proses pelatihan dan optimisasi model. Hasil transformasi tersebut dikomunikasikan secara visual dalam **Gambar 1**.



Gambar 1. Transformasi Data melalui Pengkodean Variabel Kategorikal dan Normalisasi Fitur Numerik

4.2 Ruang Pencarian Hiperparameter dan Strategi Optimisasi

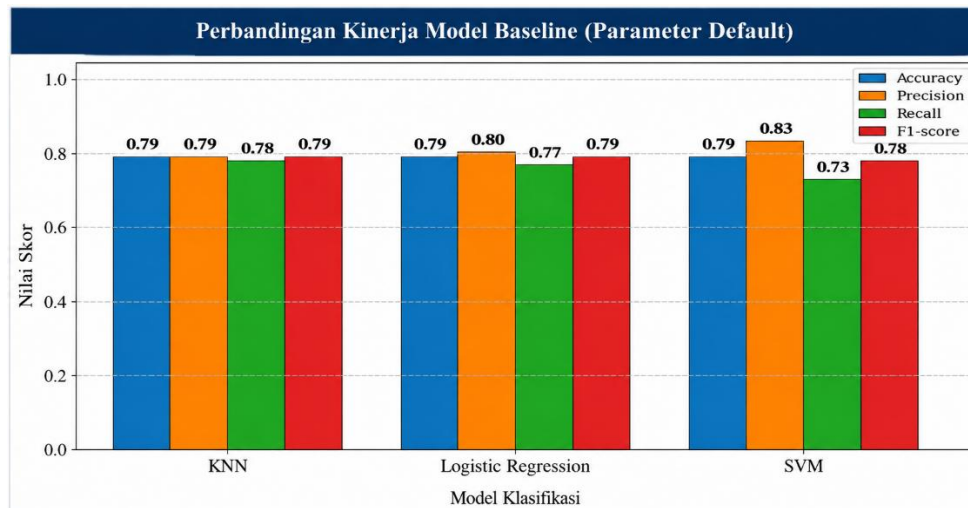
Optimisasi dilakukan menggunakan ruang pencarian yang disesuaikan dengan karakteristik masing-masing algoritma. Pada KNN, parameter yang dievaluasi meliputi *n_neighbors*, *weights*, dan *metric*; Regresi Logistik mencakup *C*, *penalty*, dan *solver*; sedangkan SVM mengevaluasi *C* dan *kernel*, dengan *gamma*='auto' dipertahankan sebagai konfigurasi tetap. Dengan demikian, *gamma* secara metodologis lebih tepat dinyatakan sebagai parameter yang dikendalikan daripada hiperparameter yang benar-benar divariasikan. Rincian fungsi dan ruang pencarian seluruh parameter dirangkum dalam **Tabel 1**, sehingga tiga tabel hiperparameter terpisah tidak lagi diperlukan.

Struktur ruang pencarian pada Tabel 1 memberikan dasar eksperimen yang konsisten bagi penerapan *Grid Search*, *Random Search*, dan *Bayesian Optimisation*. Perbandingan ketiga pendekatan tersebut tidak hanya dimaksudkan untuk menentukan konfigurasi dengan akurasi tertinggi, tetapi juga untuk mengevaluasi tingkat sensitivitas masing-masing algoritma terhadap perubahan hiperparameter. Manuskrip memang menetapkan ketiga pendekatan tersebut sebagai kerangka optimisasi komparatif.

4.3 Kinerja Model Baseline

Sebelum optimisasi, KNN, Regresi Logistik, dan SVM dievaluasi menggunakan konfigurasi parameter *default* sebagai *baseline*. Ketiga model menghasilkan **akurasi identik sebesar 0,79**, tetapi komposisi metrik lainnya memperlihatkan pola klasifikasi yang berbeda. KNN memperoleh presisi 0,79, *recall* 0,78, dan F1-score 0,79; Regresi Logistik menghasilkan presisi 0,80, *recall* 0,77, dan F1-score 0,79; sedangkan SVM mencapai presisi tertinggi sebesar **0,83**, tetapi *recall* terendah sebesar **0,73**, dengan F1-score 0,78. Hasil tersebut divisualisasikan pada **Gambar 2**.

Kesamaan akurasi sebesar 0,79 menunjukkan bahwa pemilihan model tidak dapat didasarkan hanya pada satu metrik. SVM, misalnya, menghasilkan presisi tertinggi, tetapi peningkatan tersebut diikuti penurunan *recall*. Artinya, prediksi positif SVM cenderung lebih tepat ketika model menetapkan kelas positif, tetapi proporsi observasi positif aktual yang berhasil dideteksi lebih rendah. Sebaliknya, KNN menunjukkan keseimbangan presisi dan *recall* yang lebih baik. Temuan ini mempertegas perlunya evaluasi multidimensi menggunakan akurasi, presisi, *recall*, F1-score, serta AUC-ROC apabila nilai numeriknya tersedia. Manuskrip sendiri menekankan bahwa perbedaan pola kesalahan dapat tersembunyi apabila evaluasi hanya mengandalkan akurasi.



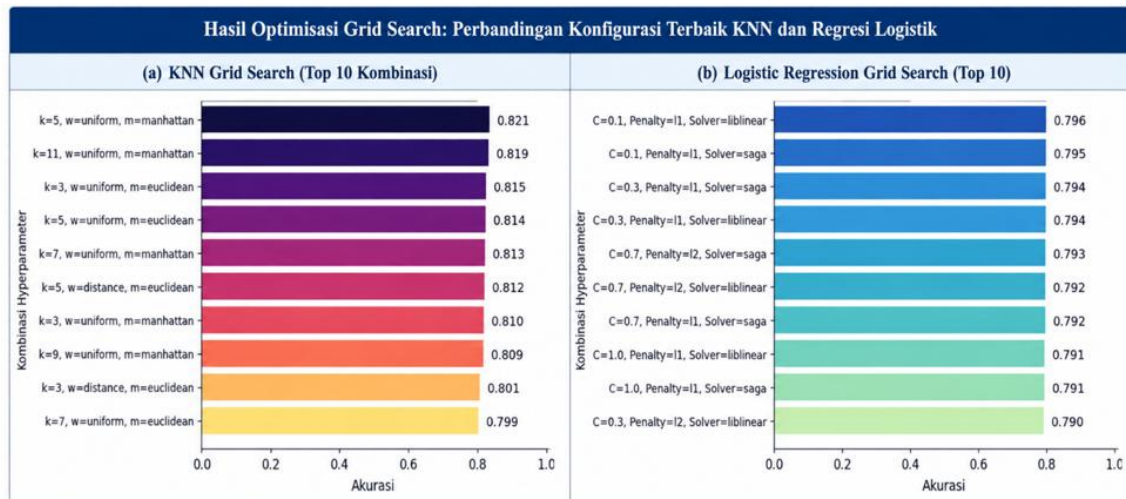
Gambar 2. Perbandingan Kinerja Model Baseline KNN, Regresi Logistik, dan SVM Berdasarkan Akurasi, Presisi, Recall, dan F1-score.

4.4 Hasil Optimisasi Menggunakan Grid Search

4.4.1 KNN dan Regresi Logistik

Hasil *Grid Search* memperlihatkan bahwa KNN merupakan model yang paling responsif terhadap optimisasi hiperparameter. Kombinasi *n_neighbors*, *weights*, dan *metric* menghasilkan akurasi sekitar 0,79–0,82. Konfigurasi berbasis *weights*=*'uniform'* dan metrik Manhattan termasuk dalam kelompok konfigurasi dengan kinerja tertinggi, dengan nilai maksimum sekitar 0,82. Dibandingkan *baseline* 0,79, peningkatan tersebut setara dengan sekitar tiga poin persentase. Perbandingan sepuluh konfigurasi terbaik KNN dan Regresi Logistik disajikan dalam **Gambar 3**.

Secara algoritmik, sensitivitas KNN dapat dijelaskan oleh ketergantungannya terhadap struktur lokal data. Perubahan jumlah tetangga dan definisi jarak secara langsung mengubah komposisi *neighbourhood* yang menjadi dasar penentuan kelas. Namun, konfigurasi dengan akurasi tertinggi tersebut tidak dapat dianggap sebagai konfigurasi universal. Keunggulannya berlaku pada ruang pencarian, transformasi fitur, dan pembagian data yang digunakan dalam eksperimen ini. Dengan demikian, kekuatan utama *Grid Search* terletak pada kemampuannya mengidentifikasi konfigurasi terbaik di dalam ruang pencarian yang ditentukan, bukan menjamin optimum global.

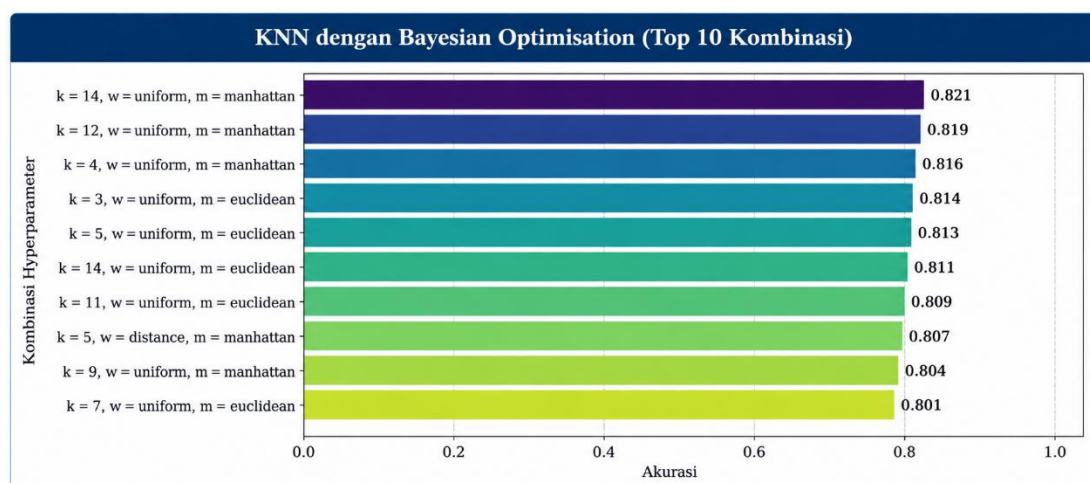


Gambar 3. Perbandingan Sepuluh Konfigurasi Hiperparameter Terbaik KNN dan Regresi Logistik Menggunakan Grid Search.

Berbeda dengan KNN, Regresi Logistik menunjukkan respons yang relatif terbatas terhadap optimisasi. Variasi C , penalty, dan solver menghasilkan akurasi yang terkonsentrasi sekitar 0,79–0,80. Peningkatan maksimum sekitar satu poin persentase dibandingkan *baseline* menunjukkan bahwa konfigurasi awal Regresi Logistik telah berada relatif dekat dengan wilayah performa terbaik dalam ruang pencarian yang diuji. Temuan ini penting karena membuktikan bahwa manfaat *hyperparameter tuning* bersifat *algorithm-dependent*, bukan peningkatan yang otomatis terjadi pada seluruh algoritma.

4.4.2 Support Vector Machine

Optimisasi SVM melalui *Grid Search* mengevaluasi variasi C dan kernel dengan $\gamma = \text{'auto'}$. Sebagian besar konfigurasi menghasilkan akurasi sekitar 0,79–0,80. Beberapa konfigurasi, termasuk $C=18$ dengan *polynomial kernel* dan $C=5$ dengan RBF, berada di antara konfigurasi dengan performa tinggi. Akan tetapi, peningkatan C tidak menunjukkan hubungan monoton dengan akurasi; nilai C yang lebih besar tidak secara otomatis menghasilkan generalisasi yang lebih baik (Gambar 4).



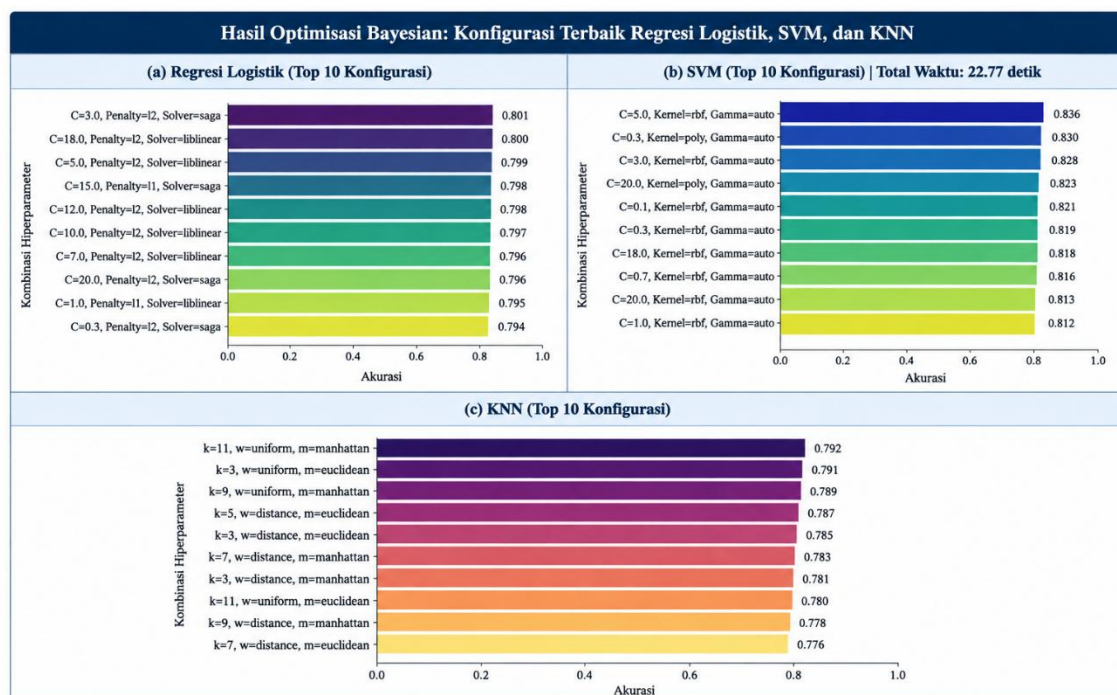
Gambar 4. Perbandingan Konfigurasi Hiperparameter Terbaik KNN dan SVM Berdasarkan Grid Search.

Temuan tersebut mengoreksi interpretasi awal yang terlalu kuat apabila SVM dinyatakan sebagai algoritma terbaik secara keseluruhan. SVM memang memiliki keunggulan presisi pada *baseline*, tetapi KNN menghasilkan akurasi maksimum yang lebih tinggi setelah optimisasi. Oleh sebab itu, bukti empiris penelitian lebih mendukung kesimpulan bahwa model terbaik bergantung pada metrik evaluasi dan tujuan klasifikasi daripada superioritas intrinsik satu algoritma.

4.5 Hasil Optimisasi Menggunakan Bayesian Optimisation

Bayesian Optimisation menghasilkan pola yang secara umum konsisten dengan *Grid Search*. Pada KNN, akurasi berada pada rentang sekitar 0,79–0,82 dan beberapa konfigurasi berbasis Manhattan mencapai nilai tertinggi sekitar **0,82**. Hasil tersebut merupakan bukti penting bahwa wilayah hiperparameter berkinerja tinggi pada KNN dapat ditemukan melalui lebih dari satu strategi pencarian. Visual KNN *Bayesian Optimisation* yang terdapat pada halaman 15 manuskrip sebaiknya tidak berdiri sendiri karena halaman yang sama juga telah menampilkan perbandingan ketiga algoritma. Saya merekomendasikan menggabungkan ketiganya menjadi satu gambar utama:

Regresi Logistik kembali menunjukkan performa yang relatif stabil pada kisaran 0,79–0,80. Beberapa kombinasi, seperti $C=3$, penalti L2 dengan solver='saga', dan $C=18$, penalti L2 dengan solver='liblinear', menghasilkan akurasi sekitar 0,80. Konsistensi hasil antara *Grid Search* dan *Bayesian Optimisation* memperkuat indikasi bahwa Regresi Logistik relatif tidak sensitif terhadap variasi hiperparameter dalam rentang yang diuji, disajikan dalam **Gambar 5**.



Gambar 5. Perbandingan Sepuluh Konfigurasi Hiperparameter Terbaik KNN, Regresi Logistik, dan SVM Menggunakan Bayesian Optimisation.

Pada SVM, beberapa konfigurasi terbaik menggunakan *kernel* RBF dan menghasilkan akurasi sekitar 0,80. Konfigurasi $C=5$, RBF, gamma='auto' dan $C=0,3$, RBF, gamma='auto'

termasuk di antara konfigurasi berkinerja tinggi. Meskipun demikian, kemunculan RBF secara berulang tidak cukup untuk menyimpulkan superioritas universal *kernel* tersebut; kesimpulan harus dibatasi pada dataset dan ruang hiperparameter yang dievaluasi.

4.6 Hasil Optimisasi Menggunakan Random Search

Berbeda dari dua pendekatan sebelumnya, *Random Search* menghasilkan performa KNN yang lebih rendah pada *search budget* yang digunakan. Sepuluh konfigurasi terbaik berada sekitar **0,77–0,79**, dengan kombinasi $k=11$, $\text{weights}=\text{'uniform'}$, dan metrik Manhattan mencapai sekitar 0,79. Dibandingkan akurasi maksimum sekitar 0,82 yang diperoleh melalui *Grid Search* dan *Bayesian Optimisation*, hasil tersebut menunjukkan bahwa *Random Search* pada konfigurasi eksperimen ini belum berhasil menemukan wilayah hiperparameter KNN dengan performa setinggi dua pendekatan lainnya.

Pada Regresi Logistik dan SVM, *Random Search* menghasilkan pola yang relatif konsisten dengan metode sebelumnya, yaitu akurasi sekitar 0,79–0,80. Regresi Logistik tetap menunjukkan perubahan terbatas terhadap variasi C , penalty , dan solver , sedangkan beberapa konfigurasi SVM berbasis RBF mencapai sekitar 0,80. Keseluruhan hasil ini memperkuat temuan bahwa sensitivitas hiperparameter bersifat spesifik algoritma: KNN menunjukkan respons terbesar, sedangkan Regresi Logistik dan SVM relatif stabil pada ruang pencarian penelitian.

4.7 Sintesis Perbandingan Model dan Metode Optimisasi

Sintesis Perbandingan Model dan Metode Optimisasi menghasilkan tiga temuan utama (Tabel 2) yaitu: **Pertama**, manfaat optimisasi paling nyata terjadi pada KNN, dengan akurasi meningkat dari 0,79 menjadi sekitar 0,82 melalui *Grid Search* maupun *Bayesian Optimisation*. Regresi Logistik dan SVM hanya mengalami peningkatan terbatas hingga sekitar 0,80. Dengan demikian, optimisasi hiperparameter terbukti bermanfaat, tetapi besar manfaatnya bergantung pada karakteristik algoritma.

Tabel 2. Perbandingan Kinerja Maksimum Model Berdasarkan Strategi Optimisasi Hiperparameter

Algoritma	Baseline	Grid Search	Bayesian Optimisation	Random Search	Interpretasi Utama
KNN	0,79	0,82	0,82	$\approx 0,79$	Paling responsif terhadap optimisasi
Regresi Logistik	0,79	$\approx 0,80$	$\approx 0,80$	$\approx 0,80$	Relatif stabil
SVM	0,79	$\approx 0,80$	$\approx 0,80$	$\approx 0,80$	Relatif stabil; RBF kompetitif

Catatan: nilai bertanda $\approx$ masih berupa angka ringkasan dalam manuskrip dan sebaiknya diganti dengan output numerik presisi dari kode eksperimen sebelum publikasi.

Kedua, bukti penelitian belum mendukung klaim bahwa *Grid Search* secara mutlak lebih unggul daripada *Bayesian Optimisation*. Keduanya menghasilkan akurasi maksimum KNN sekitar 0,82. Superioritas dari sisi efisiensi baru dapat dinilai apabila manuskrip melaporkan waktu komputasi, jumlah konfigurasi yang dievaluasi, dan *search budget* secara konsisten. Tanpa indikator tersebut, formulasi ilmiah yang lebih defensibel adalah bahwa **Grid Search dan Bayesian Optimisation menghasilkan performa maksimum yang sebanding pada KNN**, sedangkan *Random Search* menghasilkan maksimum lebih rendah pada konfigurasi eksperimen yang digunakan.

Ketiga, berdasarkan akurasi maksimum yang tersedia, **KNN hasil Grid Search atau Bayesian Optimisation merupakan kandidat model terbaik dengan akurasi sekitar 0,82**, bukan SVM sebagaimana masih dinyatakan dalam Kesimpulan versi lama. Akan tetapi, penentuan model terbaik secara definitif harus mempertimbangkan presisi, *recall*, F1-score, dan AUC-ROC hasil model teroptimisasi apabila seluruh angka tersebut tersedia. Hal ini sangat relevan karena pada *baseline*, SVM memiliki presisi tertinggi 0,83 tetapi *recall* terendah 0,73.

4.8 Implikasi, Keterbatasan, dan Robustness Temuan

Secara metodologis, hasil penelitian menegaskan bahwa optimisasi hiperparameter tidak memberikan keuntungan yang seragam bagi setiap algoritma. KNN memperoleh manfaat terbesar karena struktur keputusannya secara langsung dipengaruhi jumlah tetangga, pembobotan, dan definisi jarak. Sebaliknya, Regresi Logistik dan SVM menunjukkan wilayah performa yang relatif datar dalam ruang pencarian yang diuji. Temuan ini mengimplikasikan bahwa strategi optimisasi sebaiknya dipilih berdasarkan sensitivitas algoritma dan karakteristik data, bukan diterapkan secara mekanis dengan asumsi bahwa pencarian yang lebih kompleks selalu menghasilkan peningkatan prediktif.

Meskipun demikian, terdapat beberapa keterbatasan yang harus dinyatakan secara eksplisit. Dataset Titanic merupakan dataset historis dengan ukuran dan struktur tertentu sehingga tidak dapat secara langsung merepresentasikan risiko keselamatan maritim modern. Selain itu, selisih kinerja yang relatif kecil, misalnya peningkatan dari 0,79 menjadi 0,82, belum dengan sendirinya membuktikan perbedaan yang signifikan secara statistik. Penguatan penelitian berikutnya dapat dilakukan melalui *repeated stratified cross-validation*, estimasi interval kepercayaan, pengujian statistik antarmodel, validasi eksternal, serta pelaporan waktu komputasi dan jumlah evaluasi hiperparameter. Langkah tersebut memungkinkan efektivitas prediktif dan efisiensi komputasional ketiga strategi optimisasi dibandingkan secara lebih robust.

V. Kesimpulan

Penelitian ini menunjukkan bahwa optimisasi hiperparameter memberikan pengaruh yang berbeda terhadap kinerja K-Nearest Neighbours (KNN), Regresi Logistik (LR), dan Support Vector Machine (SVM) dalam klasifikasi kelangsungan hidup penumpang Titanic. Pada konfigurasi *baseline*, ketiga algoritma menghasilkan akurasi yang sama, yaitu 0,79, meskipun memiliki karakteristik presisi dan *recall* yang berbeda. Setelah optimisasi,

KNN menunjukkan respons paling nyata, dengan akurasi meningkat hingga sekitar 0,82 melalui *Grid Search* dan *Bayesian Optimisation*, sedangkan LR dan SVM relatif stabil pada kisaran 0,79–0,80. Temuan ini menegaskan bahwa manfaat optimisasi bersifat spesifik terhadap karakteristik algoritma dan ruang hiperparameter yang dievaluasi. *Grid Search* dan *Bayesian Optimisation* menghasilkan kinerja maksimum yang sebanding pada KNN, sementara *Random Search* belum mencapai performa yang setara dalam konfigurasi eksperimen yang digunakan. Dengan demikian, berdasarkan akurasi maksimum, KNN yang dioptimalkan melalui *Grid Search* atau *Bayesian Optimisation* merupakan konfigurasi paling kompetitif pada dataset penelitian ini. Namun, pemilihan model terbaik tetap perlu mempertimbangkan presisi, *recall*, F1-score, dan AUC-ROC secara simultan. Penelitian selanjutnya perlu menerapkan *repeated stratified cross-validation*, validasi eksternal, pengujian statistik, serta evaluasi efisiensi komputasi untuk memperkuat generalisasi dan reliabilitas model pada dataset keselamatan transportasi maritim yang lebih kompleks.

Ucapan Terimakasih

Penelitian ini terlaksana berkat pendanaan dari Kementerian Pendidikan Tinggi, Sains, dan Teknologi melalui program Penelitian Dosen Pemula (PDP) tahun 2025, sehingga penulis menyampaikan penghargaan yang setinggi-tingginya.

Daftar Pustaka

- [1] Singh, A., Saraswat, S., & Faujdar, N. (2017). Analyzing Titanic disaster using machine learning algorithms. *Proceedings of the IEEE International Conference on Computing, Communication and Automation (ICCCA)*, 406–411. <https://doi.org/10.1109/CCAA.2017.8229835>
- [2] Sudrajat, A., Faturachman, D., & Muslim, M. (2019). Analisis keselamatan transportasi penyeberangan laut dan antisipasi terhadap kecelakaan kapal di Merak–Bakauheni. *TEKNIKA: Jurnal Sains dan Teknologi*, 10(2).
- [3] Singh, J., Singh, J., Singh, G., & Kaur, N. (2022). Exploratory data analysis for interpreting model prediction using Python. *2022 International Conference on Smart Generation Computing, Communication and Networking (SMARTGENCON)*, 1–6. <https://doi.org/10.1109/SMARTGENCON56628.2022.10083533>
- [4] Satriatama, A., Wibowo, A., Arnold, I. G., Pratama, R., Masyhuda, T., Agusti, Y., Purwanti, E., & Werdiningsih, I. (2023). Analisis kluster data pasien diabetes untuk identifikasi pola dan karakteristik pasien. *Jurnal Teknologi dan Sistem Informasi Bisnis*, 5(3), 172–182. <https://doi.org/10.47233/jteksis.v5i3.828>
- [5] Nugroho, A., Wiyanto, T. S., Widiyatmoko, A. T., & Muhidin, A. (2024). Credit card fraud detection using logistic regression with hyperparameter optimization. *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*, 53–57. <https://doi.org/10.1109/ICITISEE63424.2024.10730343>
- [6] Fajri, M., & Primajaya, A. (2023). Komparasi teknik hyperparameter optimization pada SVM untuk permasalahan klasifikasi dengan menggunakan grid search dan random search. *Journal of Applied Informatics and Computing*, 7(1), 14–19. <https://doi.org/10.30871/jaic.v7i1.5004>
- [7] Laeli, A. R., Rustam, Z., Hartini, S., Maulidina, F., & Aurelia, J. E. (2020). Hyperparameter optimization on support vector machine using grid search for

- classifying thalassemia data. 2020 International Conference on Decision Aid Sciences and Applications (DASA), 817–821. <https://doi.org/10.1109/DASA51403.2020.9317227>
- [8] Gori, T., Sunyoto, A., & al Fatta, H. (2024). Preprocessing data dan klasifikasi untuk prediksi kinerja akademik siswa. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- [9] Kasanah, A. N., Muladi, M., & Pujiyanto, U. (2019). Penerapan teknik SMOTE untuk mengatasi imbalance class dalam klasifikasi objektivitas berita online menggunakan algoritma KNN. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 3(2), 196–201. <https://doi.org/10.29207/resti.v3i2.945>
- [10] Pristyanto, Y., & Wirantanu, D. (2024). Comparison of genetic algorithm and recursive feature elimination on high dimensional data. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(2), 189–198. <https://doi.org/10.29207/resti.v8i2.5375>
- [11] Shaik, M. A., Sairam, P., Rahul, P., Sadhvik, T., & Nithin, P. (2024). Titanic survival predictor using machine learning algorithms. 2024 4th Asian Conference on Innovation in Technology (ASIANCON), 1–6. <https://doi.org/10.1109/ASIANCON62057.2024.10837935>
- [12] Bisht, M., Singh, A., Tripathi, G., Shantanu, K., Gupta, A., & Gupta, R. (2024). Analysis of machine learning algorithms for predicting Titanic disaster survival rate. 2024 4th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1–6. <https://doi.org/10.1109/ICCCNT57978.2024.10352079>
- [13] Dasgupta, S., Ghosh, S., Dey, N., & Chakraborty, S. (2021). Application of machine learning techniques for Titanic dataset survival prediction. *Procedia Computer Science*, 191, 528–534. <https://doi.org/10.1016/j.procs.2021.07.068>