

Hyperparameter Optimization and Feature Selection Analysis on the XGBoost Model for Hepatitis C Infection Prediction

Nadia Martha Lefi ^{1*}, Majid Rahardi ^{2*}

* Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta
nadiamartha@students.amikom.ac.id¹, majid@amikom.ac.id²

Article Info

Article history:

Received 2025-08-22
Revised 2025-11-04
Accepted 2025-11-08

Keyword:

Classification,
Feature Selection,
Hepatitis C,
Hyperparameter Optimization,
XGBoost.

ABSTRACT

Hepatitis C is a liver disease that can progress to chronic conditions such as cirrhosis and liver cancer. Early detection is essential and can be supported through machine learning approaches. This study analyzes the effect of feature selection and hyperparameter tuning on the performance of the XGBoost model in classifying hepatitis C infection. The dataset, obtained from Kaggle, contains laboratory test attributes. The preprocessing stage involved handling missing values, encoding categorical variables, removing outlier classes, and normalizing data using StandardScaler. After stratified splitting, the training set was balanced using the SMOTE technique. Feature selection was carried out using the ANOVA F-score method, and hyperparameter tuning was performed using GridSearchCV. Three model scenarios were compared: baseline, with feature selection, and with combined feature selection and hyperparameter tuning. The evaluation results showed that the third model achieved the best performance with 96% accuracy, 79% precision, 81% recall, and a 78% F1-score, despite a slight decrease in the ROC AUC value. This approach has proven effective in improving model performance and is relevant for supporting more accurate hepatitis C diagnosis systems.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Hepatitis C adalah penyakit menular yang menyerang organ hati dan memiliki potensi untuk berkembang menjadi penyakit kronis, seperti sirosis atau kanker[1]. WHO memperkirakan bahwa pada tahun 2022 terdapat sekitar 50 juta penderita infeksi hepatitis C kronis di seluruh dunia, dengan 1 juta kasus baru per tahun[2]. Dalam dunia medis, deteksi dini infeksi hepatitis C sangat penting untuk meningkatkan angka kesembuhan, dan pembelajaran mesin mulai digunakan untuk mendukung diagnosa dengan data[3]. Extreme Gradient Boosting (XGBoost) adalah salah satu algoritma klasifikasi medis yang populer karena mampu menangani data kompleks dan menghasilkan tingkat akurasi yang tinggi[4]. Beberapa penelitian telah menggunakan XGBoost untuk memprediksi hepatitis C, tetapi belum sepenuhnya mengoptimalkan proses seleksi fitur dan tuning hyperparameter.

Penelitian ini penting karena deteksi dini Hepatitis C seringkali sulit dilakukan hanya melalui pemeriksaan manual.

Pendekatan berbasis machine learning (ML) dapat membantu mengenali pola data laboratorium yang kompleks. Tantangan utama dalam penerapan machine learning pada diagnosis medis meliputi ukuran dataset yang relatif kecil, distribusi kelas yang tidak seimbang, banyaknya fitur laboratorium, serta sensitivitas model seperti XGBoost terhadap konfigurasi hyperparameter. Dengan mempertimbangkan hal tersebut, penelitian ini mengintegrasikan seleksi fitur dan optimasi klasifikasi infeksi Hepatitis C. Dengan demikian, penelitian ini berfokus untuk mengatasi permasalahan overfitting, redundansi fitur, dan ketidakseimbangan data yang sering muncul pada model klasifikasi medis seperti XGBoost.

Penelitian sebelumnya menunjukkan bahwa meskipun XGBoost memiliki performa yang baik, model sering kali dibangun tanpa penanganan data tidak seimbang, seleksi fitur, atau tuning hyperparameter yang menyeluruh[5]. Misalnya, penelitian yang dilakukan oleh Nalasari et al. menunjukkan akurasi tinggi dengan XGBoost dan Harris Hawk Optimization, tetapi tidak mencakup proses seleksi fitur, yang menyebabkan kemungkinan overfitting[6]. Studi yang

dilakukan oleh Iparraguirre-Villanueva et al. dan Alizargar et al. hanya membandingkan algoritma tanpa melakukan optimasi model, yang menyebabkan hasil tidak stabil dan tidak dapat diterapkan secara luas[7],[8]. Selain itu, penelitian Sayadi et al. tidak mengeksplorasi model boosting seperti XGBoost secara menyeluruh dalam diagnosis hepatitis C[9]. Urgensi penerapan kedua teknik ini menjadi krusial dalam konteks diagnosis medis berbasis machine learning, karena dapat menghasilkan model yang lebih stabil, efisien, dan akurat dalam mendeteksi penyakit sejak tahap awal. Oleh karena itu, untuk mendapatkan model yang lebih kuat dan adaptif, diperlukan pendekatan sistematis yang menggabungkan seleksi fitur dan tuning hyperparameter[10].

Penggabungan seleksi fitur dan hyperparameter tuning diyakini dapat meningkatkan performa klasifikasi melalui penyederhanaan kompleksitas model dan pencarian konfigurasi optimal[11]. Seleksi fitur berperan dalam mengurangi dimensi data, meningkatkan interpretabilitas model, dan menurunkan risiko overfitting[12]. Sementara itu, hyperparameter tuning seperti GridSearchCV terbukti mampu mengoptimalkan performa model dengan menyesuaikan parameter penting, seperti learning rate dan max depth[13]. Kombinasi kedua pendekatan ini memberikan dampak signifikan terhadap ketepatan dan stabilitas prediksi, terutama dalam kasus klasifikasi medis dengan distribusi kelas tidak seimbang[14]. Dalam konteks tersebut, metode berbasis optimasi pada XGBoost menjadi solusi yang relevan untuk mendeteksi hepatitis C secara lebih akurat[15].

Penelitian ini membangun model klasifikasi infeksi hepatitis C menggunakan algoritma XGBoost berdasarkan dataset yang diperoleh dari Kaggle. Sebelum pemodelan, data dibersihkan dari nilai kosong dan diseimbangkan ulang menggunakan metode SMOTE untuk mengatasi ketidakseimbangan kelas. Seleksi fitur dilakukan dengan metode ANOVA F-score guna memilih fitur yang paling relevan. Optimasi hyperparameter diterapkan menggunakan GridSearchCV untuk memperoleh konfigurasi parameter terbaik. Evaluasi performa model dilakukan berdasarkan lima metrik utama, yaitu akurasi, precision, recall, F1-score, dan ROC AUC, dengan membandingkan tiga skenario: model dasar (baseline), model dengan seleksi fitur, serta model dengan kombinasi seleksi fitur dan hyperparameter tuning.

Hasil eksperimen menunjukkan bahwa model XGBoost dasar (baseline) menghasilkan akurasi sebesar 95%, precision 73%, recall 74%, F1-score 73%, dan ROC AUC sebesar 97%. Setelah dilakukan seleksi fitur menggunakan ANOVA F-score, nilai metrik relatif serupa dengan model baseline, meskipun terjadi sedikit penurunan pada ROC AUC menjadi 96%. Peningkatan kinerja yang lebih signifikan terlihat pada model yang menggabungkan seleksi fitur dengan hyperparameter tuning menggunakan GridSearchCV, dengan akurasi meningkat menjadi 96%, precision 79%, recall 81%, dan F1-score 78%. Namun, nilai ROC AUC pada model ini

sedikit menurun menjadi 96%, dibanding baseline. Temuan ini menunjukkan bahwa meskipun terjadi peningkatan pada sebagian besar metrik evaluasi, terdapat trade-off pada aspek kemampuan model dalam memisahkan kelas secara keseluruhan.

Penelitian ini menunjukkan bahwa penerapan hyperparameter tuning secara signifikan meningkatkan akurasi dan F1-score, yang mencerminkan keseimbangan yang lebih baik antara precision dan recall. Artinya, model menjadi lebih efektif dalam mengklasifikasikan kasus positif dan negatif dengan presisi yang lebih optimal. Meskipun terjadi sedikit penurunan pada nilai ROC AUC, hal ini mengindikasikan bahwa kemampuan model dalam membedakan antara kelas secara keseluruhan sedikit menurun, kemungkinan akibat konfigurasi parameter yang lebih agresif pasca-tuning. Namun demikian, peningkatan pada metrik utama seperti akurasi dan recall menjadikan model lebih sesuai untuk skenario klinis yang memerlukan sensitivitas tinggi. Dengan demikian, strategi gabungan antara seleksi fitur dan optimasi hyperparameter terbukti mampu mengatasi keterbatasan pendekatan sebelumnya yang belum mengintegrasikan kedua teknik tersebut secara simultan.

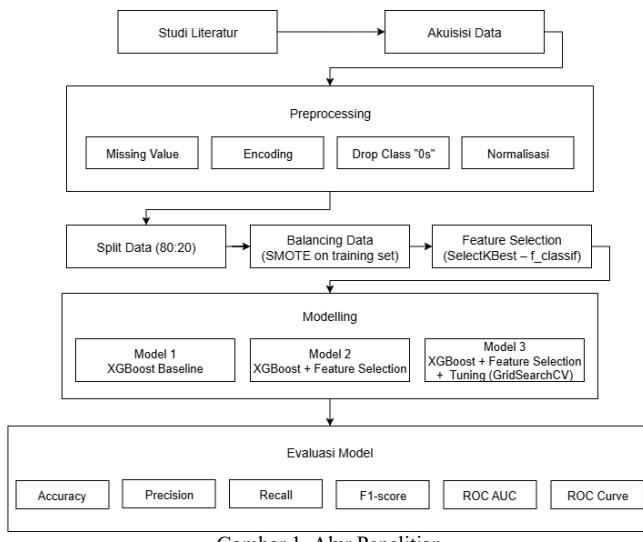
Penelitian ini menunjukkan bahwa kombinasi teknik seleksi fitur dan hyperparameter tuning pada algoritma XGBoost mampu meningkatkan kinerja model secara signifikan, terutama dalam hal akurasi, presisi, dan sensitivitas dalam klasifikasi infeksi hepatitis C. Meskipun terjadi sedikit penurunan pada nilai ROC AUC, peningkatan pada metrik lain mengindikasikan bahwa model menjadi lebih andal untuk aplikasi deteksi dini. Temuan ini menegaskan pentingnya pendekatan optimasi bertahap dalam menghasilkan model yang adaptif terhadap karakteristik data medis yang kompleks. Ke depan, pendekatan serupa dapat diterapkan pada kasus klasifikasi penyakit lain dengan distribusi kelas yang tidak seimbang dan data berdimensi tinggi. Selain itu, integrasi model ini dalam sistem pendukung keputusan klinis berpotensi mempercepat proses diagnosis dan meningkatkan kualitas pelayanan kesehatan.

Selain membandingkan tiga skenario model, penelitian ini juga mengevaluasi secara terpisah pengaruh seleksi fitur dan hyperparameter tuning terhadap peningkatan performa klasifikasi, untuk memahami kontribusi masing-masing strategi dalam konteks prediksi medis.

II. METODE

Penelitian ini dilakukan melalui pendekatan klasifikasi berbasis pembelajaran mesin menggunakan algoritma Extreme Gradient Boosting (XGBoost). Tujuan utama dari penelitian ini adalah membangun model klasifikasi untuk prediksi infeksi hepatitis C berdasarkan hasil laboratorium pasien, dengan pendekatan bertahap yang menggabungkan

teknik seleksi fitur dan tuning hyperparameter. Alur keseluruhan tahapan eksperimen ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

A. Studi Literatur

Penelitian ini dimulai dengan melakukan studi literatur terhadap berbagai sumber ilmiah seperti jurnal internasional, artikel konferensi, dan penelitian terdahulu yang relevan. Kajian difokuskan pada penerapan algoritma XGBoost untuk klasifikasi penyakit, terutama infeksi hepatitis C. Studi literatur juga mencakup pembahasan terkait metode seleksi fitur untuk meyederhanakan kompleksitas model serta teknik optimasi hyperparameter untuk meningkatkan performa klasifikasi. Selain itu, dilakukan peninjauan terhadap teknik penanganan data tidak seimbang seperti SMOTE, dan pemilihan metrik evaluasi klasifikasi multi-kelas yang tepat.

B. Akuisisi Data

Dataset penelitian diperoleh dari platform Kaggle berjudul “Hepatitis C Dataset”, yang terdiri dari 615 entri pasien dan 14 fitur, termasuk label target[16]. Atribut mencakup fitur numerik seperti Age, ALB, ALP, ALT, AST, BIL, CHE, CHOL, CREA, GGT, dan PROT, serta satu fitur kategorikal yaitu “Sex”. Label target mengindikasikan kondisi kesehatan hati dalam lima kelas, yaitu Blood Donor, Suspect Blood Donor, Hepatitis, Fibrosis, dan Cirrhosis. Dataset ini memiliki struktur yang sesuai untuk klasifikasi multi-kelas dan telah digunakan sebagai benchmark dalam sejumlah penelitian serupa.

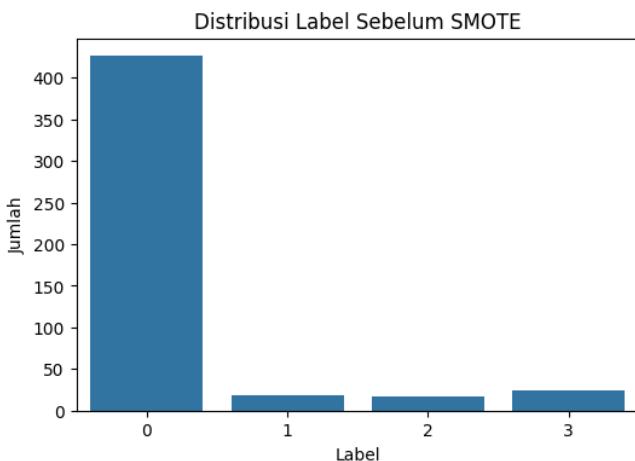
C. Preprocessing

Pada tahap preprocessing, penanganan nilai hilang dilakukan menggunakan metode K-Nearest Neighbors Imputer (KNNImputer). Metode ini menggantikan nilai kosong dengan rata-rata nilai dari tetangga terdekat dalam ruang fitur, yang terbukti efektif dalam menjaga pola distribusi data[17]. Fitur kategorikal “Sex” dikonversi menjadi nilai numerik menggunakan LabelEncoder. Kelas

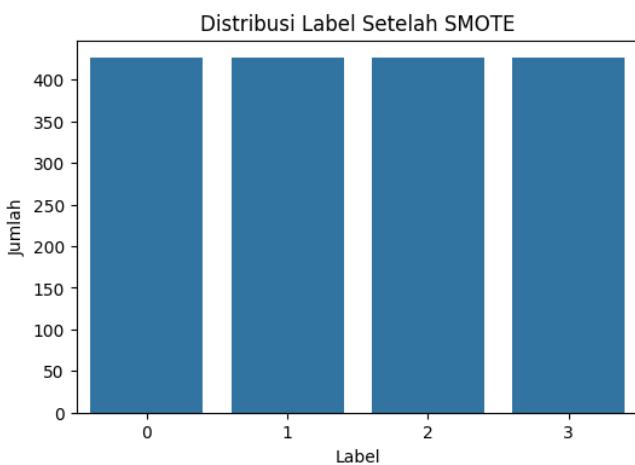
“suspect blood donor” yang hanya memiliki satu data dihapus untuk menghindari bias dalam proses pelatihan. Seluruh fitur numerik kemudian dinormalisasikan menggunakan StandardScaler, yang mengubah distribusi data menjadi skala dengan nilai rata-rata nol dan deviasi standar satu. Teknik ini penting agar model tidak terpengaruh oleh perbedaan skala antar fitur[18].

Data yang telah dibersihkan kemudian dibagi menjadi dua bagian, yakni 80% sebagai data latih dan 20% sebagai data uji. Pembagian ini dilakukan dengan metode stratifikasi agar proporsi distribusi kelas tetap terjaga di kedua subset, sesuai rekomendasi dalam studi Bushra Al-Smadi et al. untuk dataset medis tidak seimbang[19]. Tujuan pembagian ini adalah agar model dapat belajar merata dari seluruh kelas dan diuji pada data baru yang belum pernah dilihat sebelumnya.

Sebelum dilakukan SMOTE, distribusi kelas menunjukkan ketidakseimbangan yang cukup besar, yaitu 533 data pada kelas Blood Donor, 24 pada Hepatitis, 21 pada Fibrosis, dan 30 pada Cirrhosis, sementara kelas Suspect Donor hanya terdiri dari 7 data dan dihapus pada tahap preprocessing untuk menghindari bias pelatihan. Ketidakseimbangan ini menandakan dominasi kelas mayoritas (Blood Donor) yang berpotensi menyebabkan bias prediksi terhadap kelas minoritas penyakit hati. Setelah proses pembagian data menjadi data latih dan data uji, diperoleh sebanyak 486 data pada subset latih dan 122 data pada subset uji. Distribusi kelas pada data latih sebelum penerapan SMOTE masih menunjukkan ketidakseimbangan, dengan 426 data pada kelas Blood Donor, 19 pada Hepatitis, 17 pada Fibrosis, dan 24 pada Cirrhosis. Setelah dilakukan SMOTE pada data latih, jumlah total data meningkat menjadi 1.704 sampel dengan distribusi yang seimbang di setiap kelas sebesar 426 data. Penerapan teknik Synthetic Minority Over-Sampling Technique (SMOTE) dilakukan untuk mengatasi ketimpangan tersebut dengan menghasilkan data sintetis pada kelas minoritas berdasarkan interpolasi terhadap tetangga terdekat. Teknik ini hanya diterapkan pada data latih agar tidak terjadi kebocoran informasi pada data uji, serta mampu meningkatkan sensitivitas model terhadap kelas minoritas dalam klasifikasi multi kelas[20]. Proses ini memastikan bahwa model memperoleh representasi yang lebih adil dari setiap kelas sehingga dapat belajar pola secara seimbang tanpa dominasi kelas mayoritas. Dengan demikian, hasil pelatihan menjadi lebih stabil dan performa prediksi terhadap kasus penyakit hati minoritas meningkat secara signifikan. Gambar 2 menunjukkan visualisasi distribusi kelas sebelum diterapkan SMOTE, sedangkan Gambar 3 menunjukkan kondisi distribusi setelah diterapkan SMOTE.



Gambar 2. Distribusi kelas sebelum dilakukan SMOTE



Gambar 3. Distribusi kelas setelah dilakukan SMOTE

Berdasarkan kedua visualisasi tersebut, terlihat bahwa distribusi kelas menjadi lebih seimbang sehingga mendukung proses pelatihan model yang adil dan tidak bias terhadap kelas tertentu.

D. Seleksi Fitur

Setelah data seimbang, dilakukan tahap seleksi fitur menggunakan metode SelectKBest dengan fungsi evaluasi `f_classif`, yaitu ANOVA F-test. Teknik ini menghitung korelasi linier antara fitur independen dan target untuk memilih fitur paling relevan, sebagaimana juga diterapkan oleh Raufi dan Longo untuk meningkatkan efisiensi model berbasis boosting dalam sistem pengenalan pola data terstruktur[21]. Sepuluh fitur dengan nilai F tertinggi dipilih untuk digunakan pada dua skenario model. Seleksi fitur ini ditunjukkan untuk mengurangi kompleksitas model, mempercepat proses pelatihan, serta meminimalisasi overfitting.

Pemilihan metode ANOVA F-score didasarkan pada karakteristik dataset yang sebagian besar terdiri dari fitur numerik kontinu hasil pemeriksaan laboratorium. ANOVA F-score mampu mengevaluasi tingkat relevansi setiap fitur

terhadap variabel target berdasarkan perbedaan rata-rata antar kelas, sehingga efisien dan mudah diinterpretasikan untuk data medis. Metode alternatif seperti Mutual Information atau Boruta tidak digunakan karena memiliki kompleksitas komputasi lebih tinggi dan memerlukan waktu pelatihan yang lebih lama, sedangkan ukuran dataset penelitian ini relatif kecil. Oleh karena itu, ANOVA F-score dipilih untuk menjaga keseimbangan antara efisiensi dan ketepatan dalam seleksi fitur.

E. Modelling dan Optimasi Hyperparameter

Model klasifikasi dikembangkan dengan algoritma XGBoost yang berbasis pohon keputusan dan terkenal akan efisiensinya dalam menangani data tabular. Tiga skenario model diterapkan dalam penelitian ini. Model pertama merupakan baseline tanpa seleksi fitur maupun tuning parameter. Model kedua dibangun dengan data hasil seleksi fitur. Model ketiga menggabungkan seleksi fitur dengan optimasi hyperparameter menggunakan GridSearchCV. Pendekatan ini memungkinkan pencarian parameter terbaik seperti `n_estimators`, `max_depth`, `learning_rate`, dan `subsample` melalui teknik pencarian grid dan validasi silang sebanyak lima kali. Konfigurasi terbaik kemudian digunakan untuk membangun ulang model dengan kinerja optimal.

GridSearchCV dipilih karena metode ini mampu melakukan pencarian kombinasi parameter secara menyeluruh dan sistematis dengan evaluasi kinerja melalui cross-validation, sehingga dapat menemukan konfigurasi parameter terbaik yang meningkatkan generalisasi model. Dalam penelitian ini, parameter yang dioptimasi meliputi `n_estimators` ([100, 200]), `max_depth` ([3, 5]), `learning_rate` ([0.01, 0.1]), dan `subsample` ([0.8, 1.0])

Keempat parameter tersebut berpengaruh terhadap kompleksitas dan stabilitas model XGBoost, dimana `learning_rate` mengontrol kecepatan pembelajaran, `max_depth` mengatur kedalaman pohon keputusan, `n_estimators` menentukan jumlah pohon, sedangkan `subsample` mengontrol proporsi data yang digunakan untuk setiap pohon agar mengurangi risiko overfitting. Proses tuning dilakukan dengan 5-fold cross-validation menggunakan metrik evaluasi F1-macro, yang memastikan performa model optimal di seluruh kelas tanpa bias terhadap kelas mayoritas.

F. Evaluasi Model

Evaluasi dilakukan dengan lima metrik utama, yaitu akurasi, precision, recall, F1-score, dan ROC AUC dengan pendekatan macro-averaging. Macro average digunakan untuk memastikan kontribusi setiap kelas dinilai secara setara. Visualisasi berupa kurva ROC multi-kelas ditampilkan untuk mengamati kemampuan model dalam membedakan antar kelas. Rumus evaluasi untuk masing-masing metrik ditunjukkan pada Persamaan (1) hingga (5) :

$$\text{accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (5)$$

Ketiga skenario model dievaluasi secara menyeluruh, dengan perbandingan hasil masing-masing metrik untuk mengukur sejauh mana pengaruh seleksi fitur dan hyperparameter tuning terhadap kinerja model dalam prediksi infeksi hepatitis C. TABEL I merangkum skema evaluasi dari masing-masing model.

TABEL I
SKEMA EVALUASI MODEL

Model	Konfigurasi dan Evaluasi Model		
	Seleksi Fitur	Tuning Hyperparameter	Metrik Evaluasi
XGBoost Baseline	Tidak	Tidak	Akurasi, Precision, Recall, F1, AUC
XGBoost + Feature Selection	Ya	Tidak	Akurasi, Precision, Recall, F1, AUC
XGBoost + FS + Tuning	Ya	Ya	Akurasi, Precision, Recall, F1, AUC

Eksperimen dilaksanakan menggunakan Google Colab berbasis Python 3.10 dengan pustaka utama scikit-learn, xgboost, imbalanced-learn, numpy, dan pandas. Eksekusi dilakukan pada perangkat laptop dengan processor Intel Core i3, RAM 16 GB.

III. HASIL DAN PEMBAHASAN

A. Hasil Evaluasi Model

Evaluasi model dilakukan untuk mengukur efektivitas algoritma XGBoost dalam mengklasifikasi infeksi hepatitis C berdasarkan hasil laboratorium pasien. Tiga skenario diuji secara bertahap, yaitu model dasar tanpa optimasi, model dengan seleksi fitur, dan model dengan kombinasi seleksi fitur serta hyperparameter tuning menggunakan GridSearchCV. Evaluasi dilakukan menggunakan lima metrik utama, yaitu akurasi, precision, recall, F1-score, dan ROC AUC dengan pendekatan macro-average.

Hasil evaluasi menunjukkan bahwa model baseline memperoleh akurasi sebesar 95%, precision 73%, recall 74%, F1-score 73%, dan ROC AUC 97%. Setelah dilakukan seleksi fitur menggunakan metode ANOVA F-score, performa model

tidak berubah secara signifikan, kecuali sedikit penurunan pada ROC AUC menjadi 96%. Peningkatan performa yang paling terlihat terjadi pada model dengan kombinasi seleksi fitur serta hyperparameter tuning, yang menghasilkan akurasi 96%, precision 79%, recall 81%, dan F1-score 78%, meskipun nilai ROC AUC mengalami sedikit penurunan menjadi 96%. TABEL II menampilkan ringkasan hasil evaluasi dari ketiga skenario model.

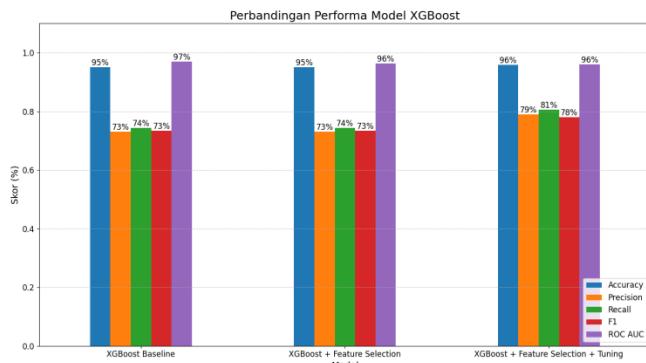
TABEL II
METRIK EVALUASI MODEL

Model	Akurasi	Precision	Recall	F1-score	ROC AUC
XGBoost Baseline	95%	73%	74%	73%	97%
XGBoost + Feature Selection	95%	73%	74%	73%	96%
XGBoost + FS + Tuning	96%	79%	81%	78%	96%

Meskipun perbedaan performa antara ketiga skenario model belum diuji secara statistik menggunakan metode seperti paired t-test karena keterbatasan ukuran dataset, konsistensi peningkatan pada sebagian besar metrik, khususnya akurasi, precision, recall, dan F1-score, menunjukkan adanya peningkatan yang signifikan secara praktis. Hal ini menandakan bahwa kombinasi feature selection dan hyperparameter tuning memberikan dampak nyata terhadap kemampuan model dalam melakukan klasifikasi infeksi Hepatitis C.

B. Visualisasi Hasil Evaluasi

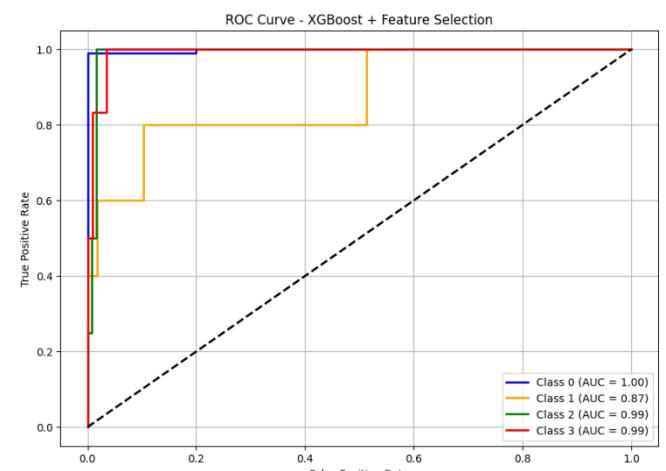
Dalam bagian ini, untuk memperjelas pola perubahan performa antar model, Gambar 4 menunjukkan grafik perbandingan metrik evaluasi dari masing-masing model. Terlihat bahwa akurasi terus meningkat dari model baseline hingga model tuning, sementara ROC AUC justru sedikit menurun pada model terakhir. Grafik ini memperkuat hasil numerik pada TABEL II dan mengilustrasikan trade-off antara sensitivitas dan pemisahan kelas global. Perbedaan tren ini menunjukkan bahwa peningkatan performa model lebih menekankan pada sensitivitas terhadap kasus positif dibandingkan pemisahan kelas secara keseluruhan.



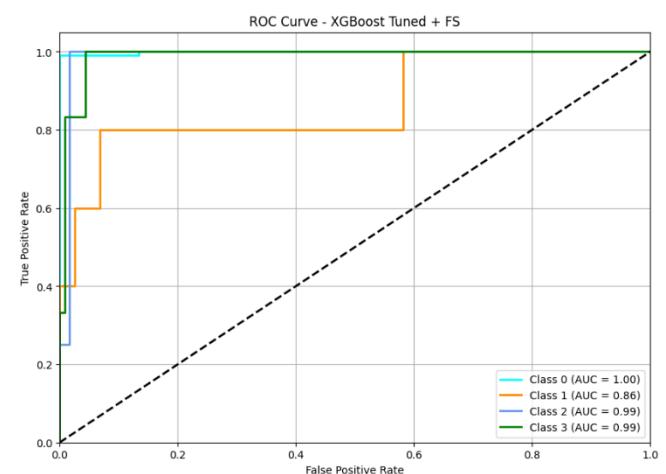
Gambar 4. Perbandingan Performa Model

Hal ini menunjukkan bahwa hyperparameter tuning mampu meningkatkan kemampuan model dalam mendeteksi kasus positif (kelas minoritas), meskipun sedikit mengorbankan kemampuan membedakan antar kelas secara keseluruhan (AUC).

Visualisasi ROC Curve pada Gambar 5, Gambar 6, dan Gambar 7 menunjukkan perbandingan kemampuan masing-masing model dalam membedakan antar kelas. Ketiga kurva memperlihatkan performa yang tinggi secara keseluruhan, namun model baseline memiliki AUC yang sedikit lebih luas, mengindikasikan kemampuan pemisahan antar kelas yang lebih baik secara global. Meski demikian, model dengan seleksi fitur dan tuning menunjukkan peningkatan stabilitas pada area true positive rate, yang berarti lebih sensitif terhadap deteksi kasus positif. Penurunan kecil pada nilai AUC menggambarkan adanya trade-off antara kemampuan pemisahan kelas dan peningkatan sensitivitas model terhadap kelas minoritas. Secara keseluruhan, hasil ini menunjukkan bahwa proses optimasi tidak hanya meningkatkan performa metrik utama, tetapi juga membuat model lebih adaptif terhadap distribusi data yang tidak seimbang.



Gambar 6. ROC Curve Model 2

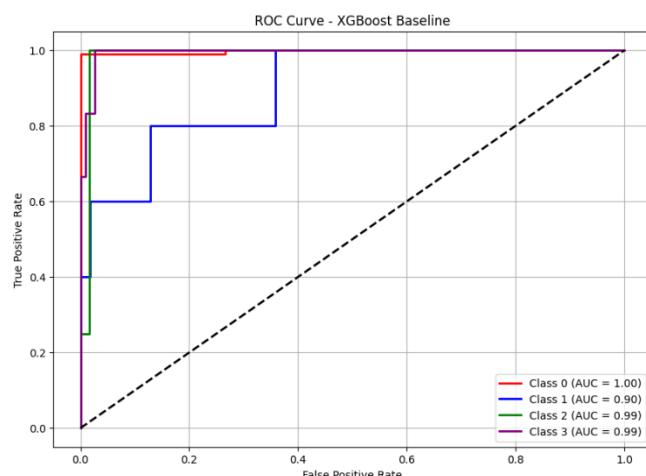


Gambar 7. ROC Curve Model 3

C. Analisis Per Model

Model dengan kombinasi seleksi fitur dan hyperparameter tuning menunjukkan peningkatan pada akurasi, precision, recall, dan F1-score. Akurasi yang lebih tinggi mencerminkan kemampuan model mengklasifikasikan data secara keseluruhan dengan lebih tepat, sementara precision yang lebih baik menunjukkan efektivitas dalam menghindari false positive. Nilai recall yang lebih tinggi mengindikasikan kemampuan deteksi kasus positif yang lebih kuat, dan peningkatan F1-score menandakan keseimbangan yang baik antara presisi dan sensitivitas yang sangat penting dalam klasifikasi medis.

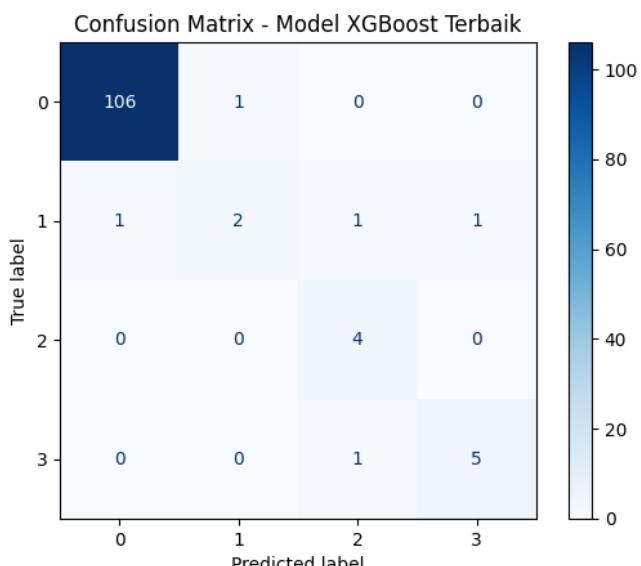
Peningkatan performa setelah dilakukan feature selection dan hyperparameter tuning menunjukkan bahwa kedua proses tersebut berperan penting dalam mengurangi noise dan meningkatkan kemampuan generalisasi model. Seleksi fitur dengan metode ANOVA F-score membantu mengeliminasi atribut laboratorium yang kurang relevan, sehingga model berfokus hanya pada fitur yang benar-benar berkontribusi terhadap prediksi infeksi hepatitis C. Hal ini membuat model lebih efisien dan mengurangi risiko overfitting.



Gambar 5. ROC Curve Model 1

Sementara itu, optimasi hyperparameter melalui GridSearchCV menyesuaikan parameter penting seperti learning_rate, max_depth, n_estimators, dan subsample, yang berperan dalam mengontrol kompleksitas model. Penyesuaian parameter ini menghasilkan keseimbangan antara bias dan variansi, sehingga performa meningkat pada sebagian metrik evaluasi.

Untuk memperjelas hasil klasifikasi, Gambar 8 menampilkan Confusion Matrix dari model terbaik, yaitu XGBoost + Feature Selection + Hyperparameter Tuning. Mayoritas data berhasil diklasifikasikan dengan benar, terutama pada kelas 0 (non-hepatitis) dengan jumlah prediksi benar sebanyak 106 sampel. Sebagian kecil kesalahan klasifikasi terjadi pada kelas minoritas, seperti hepatitis, fibrosis, dan cirrhosis, yang masing-masing memiliki jumlah data lebih sedikit dibanding kelas utama.



Gambar 8. Confusion Matrix

Berdasarkan confusion matrix tersebut, model menunjukkan kemampuan deteksi yang baik terhadap kasus positif (kelas minoritas), dengan jumlah false negative yang relatif rendah. Dalam konteks medis, hal ini menunjukkan bahwa model dapat berfungsi sebagai alat bantu skrining awal untuk mendeteksi potensi infeksi Hepatitis C berdasarkan hasil laboratorium. Model ini juga berpotensi diintegrasikan dalam Clinical Decision Support System (CDSS) untuk membantu tenaga kesehatan dalam diagnosis awal secara cepat dan akurat, terutama pada kasus dengan data laboratorium yang kompleks.

Penurunan kecil pada nilai ROC AUC pada model ketiga menunjukkan adanya trade-off antara peningkatan sensitivitas akibat konfigurasi parameter yang lebih eksploratif, dengan kemampuan model dalam membedakan antar kelas secara keseluruhan. Meskipun demikian, peningkatan pada precision dan recall memberikan keunggulan praktis yang lebih besar, terutama dalam konteks aplikasi klinis yang memerlukan

sensitivitas tinggi. Hasil ini sejalan dengan studi oleh Christopher Meaney et al. yang menyatakan bahwa tuning hyperparameter pada XGBoost dapat meningkatkan sensitivitas, namun sering diikuti oleh penurunan AUC akibat konfigurasi parameter yang lebih kompleks pada dataset kesehatan berimbang[22].

Secara keseluruhan, pendekatan bertahap dengan seleksi fitur dan optimasi hyperparameter terbukti memberikan hasil yang lebih baik dibanding model dasar. Strategi pengembangan model yang sistematis ini dapat meningkatkan akurasi prediksi, khususnya dalam klasifikasi multi-kelas dengan distribusi tidak seimbang seperti pada kasus infeksi hepatitis C. Meskipun terjadi sedikit penurunan pada nilai ROC AUC, peningkatan signifikan pada metrik lainnya menunjukkan bahwa model ketiga lebih andal untuk deteksi hepatitis C secara akurat.

IV. KESIMPULAN

Penelitian ini menegaskan bahwa pendekatan bertahap dalam membangun model klasifikasi yang menggabungkan seleksi fitur dan optimasi hyperparameter pada algoritma XGBoost, mampu meningkatkan efektivitas deteksi dini infeksi hepatitis C. Dengan mengintegrasikan proses pembersihan data, penanganan ketidakseimbangan kelas, pemilihan fitur relevan, serta pencarian konfigurasi parameter optimal, model menjadi lebih adaptif terhadap kompleksitas data medis.

Hasil evaluasi menunjukkan bahwa model XGBoost baseline memperoleh akurasi 95%, precision 73%, recall 74%, F1-score 73%, dan ROC AUC 97%. Setelah dilakukan seleksi fitur, performa model relative serupa dengan baseline, dengan ROC AUC sedikit menurun menjadi 96%. Peningkatan signifikan diperoleh pada model yang menggabungkan seleksi fitur dan hyperparameter tuning, dengan akurasi meningkat menjadi 96%, precision 79%, recall 81%, dan F1-score 78%, meskipun nilai ROC AUC tetap pada kisaran 96%.

Peningkatan performa ini menunjukkan bahwa kombinasi seleksi fitur dan optimasi parameter membantu mengurangi noise pada data serta memperkuat generalisasi model. Meskipun terjadi sedikit penurunan nilai ROC AUC, hal tersebut merupakan konsekuensi dari peningkatan sensitivitas terhadap kelas minoritas. Model menjadi lebih baik dalam mendeteksi kasus positif Hepatitis C, namun sedikit menurunkan kemampuan pemisahan antar kelas secara global.

Strategi pengembangan model yang sistematis ini tidak hanya meningkatkan performa prediktif, tetapi juga memperkuat pemanfaatan machine learning dalam sistem pendukung keputusan klinis, khususnya pada tahap screening awal berbasis data laboratorium. Pendekatan ini menunjukkan potensi implementasi model XGBoost yang telah dioptimasi sebagai alat bantu deteksi dini dalam konteks medis.

Meskipun memberikan performa yang terbaik, penelitian ini memiliki beberapa keterbatasan, yaitu ukuran dataset yang

relatif kecil, penggunaan satu sumber data (Kaggle), tidak adanya validasi eksternal lintas dataset, dan fokus hanya pada satu algoritma (XGBoost). Ke depan, pendekatan serupa dapat dieksplorasi lebih lanjut untuk kasus klasifikasi penyakit lain, terutama yang memiliki distribusi kelas tidak seimbang dan banyak fitur input. Selain itu, penelitian selanjutnya disarankan untuk menerapkan metode ini pada dataset medis yang lebih besar dan bervariasi, membandingkan dengan algoritma lain seperti Random Forest atau LightGBM, serta mengintegrasikan model ke dalam sistem klinis real-time untuk mendukung proses deteksi dini infeksi secara otomatis dan efisien.

Secara ilmiah, hasil penelitian ini menunjukkan bahwa kombinasi antara seleksi fitur dan optimasi hyperparameter tidak hanya meningkatkan performa numerik, tetapi juga memperkuat interpretabilitas hasil model dalam konteks medis. Fitur-fitur yang terpilih melalui ANOVA F-score, seperti ALT, AST, dan BIL, merupakan indikator utama fungsi hati yang relevan terhadap deteksi infeksi Hepatitis C. Dengan demikian, peningkatan performa model dapat diartikan bukan sekadar hasil dari optimasi algoritmik, melainkan juga karena pemodelan yang lebih selaras dengan karakteristik biologis penyakit.

DAFTAR PUSTAKA

- [1] C. Shen, X. Jiang, M. Li, and Y. Luo, "Hepatitis Virus and Hepatocellular Carcinoma: Recent Advances," *Cancers (Basel)*, vol. 15, no. 2, 2023, doi: 10.3390/cancers15020533.
- [2] World Health Organization., *Global Hepatitis Report 2024*. 2024. [Online]. Available: <https://www.who.int/publications/item/9789240091672>
- [3] A. Sharma, T. Khade, and S. M. Satapathy, "A cross dataset meta-model for hepatitis C detection using multi-dimensional pre-clustering," *Sci Rep*, vol. 15, no. 1, pp. 1–17, 2025, doi: 10.1038/s41598-025-91298-0.
- [4] Y. Wang, B. Yin, and Q. Zhu, "Application of Machine Learning Algorithms in Predicting Hepatitis C," *ACM International Conference Proceeding Series*, pp. 359–365, 2023, doi: 10.1145/3644116.3644176.
- [5] A. Zulfiqar, T. Iqbal, and A. Munir, "Machine Learning-Based Classification Algorithms for Predicting Hepatitis C: A Comprehensive Analysis," *JCBI*, vol. 08, no. 01, 2024.
- [6] L. T. Nalasari, S. Anam, and N. Shofianah, "Liver Cirrhosis Classification using Extreme Gradient Boosting Classifier and Harris Hawk Optimization as Hyperparameter Tuning," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 2, pp. 508–519, 2025, doi: 10.35882/jeeemi.v7i2.730.
- [7] O. Iparraguirre-villanueva, R. O. Flores-castañeda, H. Cherrovaldivieso, and O. Iparraguirre-villanueva, "Predicting hepatitis C infection with machine learning algorithms: a prospective study," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 4, pp. 4403–4413, 2024, doi: 10.11591/ijai.v13.i4.pp4403-4413.
- [8] A. Alizargar, Y. L. Chang, and T. H. Tan, "Performance Comparison of Machine Learning Approaches on Hepatitis C Prediction Employing Data Mining Techniques," *Bioengineering*, vol. 10, no. 4, 2023, doi: 10.3390/bioengineering10040481.
- [9] M. Sayadi, V. Varadarajan, E. Gozali, and M. Sadeghi, "Effective factors in diagnosing the degree of hepatitis C using machine learning," *Frontiers in Health Informatics*, vol. 12, 2023, doi: 10.30699/fhi.v12i0.440.
- [10] B. Bischi *et al.*, "Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges," *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 13, no. 2, 2023, doi: 10.1002/widm.1484.
- [11] K. Cao *et al.*, "Prediction of cardiovascular disease based on multiple feature selection and improved PSO-XGBoost model," *Sci Rep*, vol. 15, no. 1, pp. 1–12, 2025, doi: 10.1038/s41598-025-96520-7.
- [12] A. M. Ali *et al.*, "Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection," *Machines*, vol. 11, no. 3, pp. 1–14, 2023, doi: 10.3390/machines11030391.
- [13] A. A. Syahputra and R. E. Saputro, "Application of the XGBoost Model with Hyperparameter Tuning for Industry Classification for Job Applicants," *Sinkron*, vol. 8, no. 3, pp. 1920–1931, 2024, doi: 10.3395/sinkron.v8i3.13840.
- [14] E. Pérez-gómez *et al.*, "Exploratory integration of near-infrared spectroscopy with clinical data: a machine learning approach for HCV detection in serum samples," *Front Med (Lausanne)*, 2025.
- [15] M. Wiens, A. Verone-Boyle, N. Henscheid, J. T. Podichetty, and J. Burton, "A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) Artificial Intelligence Algorithm for Drug Development Applications," *Clin Transl Sci*, vol. 18, no. 3, p. e70172, 2025, doi: 10.1111/cts.70172.
- [16] Wearefuture01, "Hepatitis C Prediction." [Online]. Available: <https://www.kaggle.com/datasets/wearefuture01/hepatitis-c-prediction>
- [17] A. Juna *et al.*, "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron," *Water (Switzerland)*, vol. 14, no. 17, pp. 1–19, 2022, doi: 10.3390/w14172592.
- [18] Q. A. Hidayaturrohman and E. Hanada, "Impact of Data Pre-Processing Techniques on XGBoost Model Performance for Predicting All-Cause Readmission and Mortality Among Patients with Heart Failure," *BioMedInformatics*, vol. 4, no. 4, pp. 2201–2212, 2024, doi: 10.3390/biomedinformatics4040118.
- [19] B. Al-Smadi, B. Hammou, H. Faris, and P. A. Castillo, "Enhancing the Classification of Imbalanced Arabic Medical Questions Using DeepSMOTE," *AI (Switzerland)*, vol. 6, no. 4, pp. 1–26, 2025, doi: 10.3390/ai6040077.
- [20] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective Class-Imbalance Learning Based on SMOTE and Convolutional Neural Networks," *Applied Sciences (Switzerland)*, vol. 13, no. 6, 2023, doi: 10.3390/app13064006.
- [21] B. Raufi and L. Longo, "Comparing ANOVA and PowerShap Feature Selection Methods via Shapley Additive Explanations of Models of Mental Workload Built with the Theta and Alpha EEG Band Ratios," *BioMedInformatics*, vol. 4, no. 1, pp. 853–876, 2024, doi: 10.3390/biomedinformatics4010048.
- [22] C. Meaney, X. Wang, J. Guan, and T. A. Stukel, "Comparison of methods for tuning machine learning model hyper-parameters: with application to predicting high-need high-cost health care users," *BMC Med Res Methodol*, vol. 25, no. 1, 2025, doi: 10.1186/s12874-025-02561-x.