# Perbandingan Metode Seleksi Fitur *Filter* dan *Wrapper* pada Klasifikasi Risiko Penyakit Jantung menggunakan *k*-NN

## *Comparison of Filter and Wrapper Feature Selection Methods for Heart Disease Risk Classification using K-Nearest Neighbors (k-NN)*

**[1]Deni Kuswandani*, [2]Herman, [3]Rusydi Umar**
[1,2]Program Studi Magister Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, 55191
[3]Program Studi Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, 55191
[1,2,3]Jl. Prof. DR. Soepomo Sh, Kota Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
*e-mail: 2408048022@webmail.uad.ac.id

(***received*:** 22 December 2025, ***revised*:** 13 January 2026, ***accepted*:** 14 January 2026)

**Abstrak**

Pemilihan metode seleksi fitur yang tepat berperan penting dalam meningkatkan efektivitas model klasifikasi medis. Penelitian ini membandingkan dua pendekatan seleksi fitur, yaitu metode *filter* dan *wrapper*, dalam membangun model *k-Nearest Neighbors* (k-NN) untuk klasifikasi risiko penyakit jantung. Dataset yang digunakan terdiri dari data demografis, gaya hidup, dan indikator klinis pasien. Pada penelitian ini, metode *filter* diterapkan dengan menyesuaikan tipe data. *Pearson Correlation* digunakan untuk fitur numerik, sedangkan *Chi-Square Test* digunakan untuk fitur kategorikal. Hasil seleksi dari kedua pendekatan ini kemudian digabungkan, sehingga dari total 20 fitur awal hanya tersisa empat variabel utama yang dinilai paling relevan terhadap klasifikasi risiko penyakit jantung, yaitu *BMI*, *Homocysteine Level*, *Blood Pressure* dan *Stress Level*, menghasilkan efisiensi komputasi tinggi namun peningkatan akurasi yang minimal (76,8%) dan *recall* kelas minoritas yang rendah (0,07). Sebaliknya, metode *wrapper* dengan *Sequential Forward Selection* (SFS) menghasilkan subset 11 fitur yang lebih informatif, dengan akurasi meningkat hingga 80,00% dan ROC-AUC mencapai 0,657, menunjukkan kemampuan diskriminasi yang lebih baik terhadap kelas minoritas. Hasil ini menunjukkan bahwa metode *filter* unggul dalam kesederhanaan dan kecepatan, sedangkan metode *wrapper* lebih efektif dalam meningkatkan performa klasifikasi. Temuan ini memberikan kontribusi empiris terhadap pemilihan strategi seleksi fitur yang sesuai dengan tujuan analisis dalam sistem pendukung keputusan klinis.

**Kata kunci:** *filter method*, *k-Nearest Neighbors* (*k*-NN), klasifikasi penyakit jantung, seleksi fitur, *wrapper method*.

***Abstract***

*Feature selection plays a crucial role in improving the effectiveness of medical classification models. This study compares two feature selection approaches—filter and wrapper methods—in developing a k-Nearest Neighbors (k-NN) model for heart disease risk classification. The dataset consists of patients' demographic data, lifestyle factors, and clinical indicators. In this study, the filter method was applied by considering data types: Pearson Correlation was used for numerical features, while the Chi-Square test was applied to categorical features. The selected features from both techniques were then combined, reducing the initial 20 features to four key variables considered most relevant for heart disease risk classification: BMI, homocysteine level, blood pressure, and stress level. This approach achieved high computational efficiency; however, it resulted in only a modest accuracy improvement (76.8%) and a low recall for the minority class (0.07). In contrast, the wrapper method using Sequential Forward Selection (SFS) produced a more informative subset of 11 features, achieving higher accuracy (80.00%) and a ROC-AUC of 0.657, indicating better discrimination capability for the minority class. These findings suggest that while the filter method excels in simplicity and computational efficiency, the wrapper method is more effective in improving classification performance. This study provides empirical insights into selecting appropriate feature selection strategies based on analytical objectives, particularly for clinical decision support systems.*

 ISSN:2302-8149
e-ISSN:2540-9719

***Keywords:*** *feature selection, filter method, heart disease classification, k-Nearest Neighbors (k-NN), wrapper method.*

## 1 Pendahuluan

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia, Menurut[1] penyakit kardiovaskular merupakan penyebab utama kematian secara global, dengan sekitar 17,9 juta jiwa meninggal setiap tahun, setara 32% dari seluruh kematian di dunia, dan 85% di antaranya disebabkan oleh serangan jantung serta *stroke*. Deteksi dini terhadap risiko penyakit jantung sangat penting untuk membantu proses pencegahan dan perawatan. Dengan berkembangnya teknologi *machine learning*, analisis data kesehatan menjadi semakin relevan untuk menghasilkan prediksi risiko yang lebih akurat.

Salah satu tantangan dalam penerapan *machine learning* adalah keberadaan fitur yang tidak relevan atau redundan. Fitur-fitur tersebut dapat menurunkan kinerja model, menambah kompleksitas komputasi, serta meningkatkan kemungkinan *overfitting* [2] [3]. Oleh karena itu, dibutuhkan teknik seleksi fitur untuk memilih variabel yang paling relevan terhadap target.

Berbagai pendekatan telah dikembangkan untuk melakukan seleksi fitur, diantaranya *filter method* yang bersifat sederhana dan efisien, serta *wrapper method* yang mengevaluasi kombinasi fitur secara langsung dengan model klasifikasi, keduanya terbukti dapat meningkatkan performa model klasifikasi medis. Namun, sebagian besar penelitian hanya menitikberatkan pada satu pendekatan (*filter* atau *wrapper*) atau satu algoritma klasifikasi. Selain itu, mayoritas studi masih mengutamakan akurasi sebagai metrik utama, tanpa memperhatikan ROC-AUC atau isu ketidakseimbangan data yang umum terjadi dalam dataset medis. Kajian spesifik pada klasifikasi penyakit jantung berbasis *k-Nearest Neighbors* (k-NN) dengan perbandingan langsung antara metode *filter* dan *wrapper* juga masih terbatas.

Dengan demikian, penelitian ini menjadi salah satu studi yang secara eksplisit membandingkan metode seleksi fitur *filter* dan *wrapper* pada klasifikasi penyakit jantung berbasis k-NN dalam kondisi dataset tidak seimbang (*imbalanced*) dengan evaluasi ROC-AUC serta recall kelas minoritas.

Berdasarkan gap tersebut, penelitian ini dirumuskan untuk menjawab beberapa pertanyaan riset utama. Pertama, penelitian ini mengkaji bagaimana perbandingan performa metode seleksi fitur *filter* dan *wrapper* terhadap klasifikasi risiko penyakit jantung menggunakan algoritma *k-Nearest Neighbors* (k-NN). Kedua, penelitian ini menganalisis apakah *wrapper method*, khususnya *Sequential Forward Selection*, mampu meningkatkan nilai ROC-AUC serta recall kelas minoritas dibandingkan *filter method* pada kondisi data yang tidak seimbang (*imbalanced*). Ketiga, penelitian ini juga bertujuan untuk mengidentifikasi fitur-fitur yang paling relevan dalam mendukung prediksi risiko penyakit jantung berdasarkan metode seleksi fitur yang digunakan.

## 2 Tinjauan Literatur

Seleksi fitur telah banyak digunakan untuk meningkatkan performa klasifikasi pada domain kesehatan dengan menyederhanakan model dan mengurangi fitur redundan. Pendekatan *filter* populer karena sederhana, cepat, dan *model-agnostic*. Misalnya, [4] dan [5] menunjukkan *filter* mampu meningkatkan efisiensi dan akurasi, serta [6] menegaskan penggunaannya sebagai langkah awal dalam prediksi penyakit jantung. Namun, secara metodologis *filter method* mengevaluasi fitur secara univariat, sehingga mengabaikan interaksi antar variabel klinis yang bersifat multivariat dan saling bergantung pada kasus penyakit jantung.

Sebaliknya, *wrapper method* mengevaluasi subset fitur langsung melalui algoritma klasifikasi tertentu. Studi [7] dan [8] membuktikan *wrapper method* dapat menghasilkan performa lebih optimal karena mempertimbangkan kombinasi fitur, di mana secara metodologis pendekatan ini mampu menangkap interaksi klinis antar variabel, meskipun dengan konsekuensi peningkatan kompleksitas dan biaya komputasi. Kombinasi *filter method* dan *wrapper method* juga dilaporkan meningkatkan kinerja pada klasifikasi penyakit jantung dan kasus medis lain [9] [10].

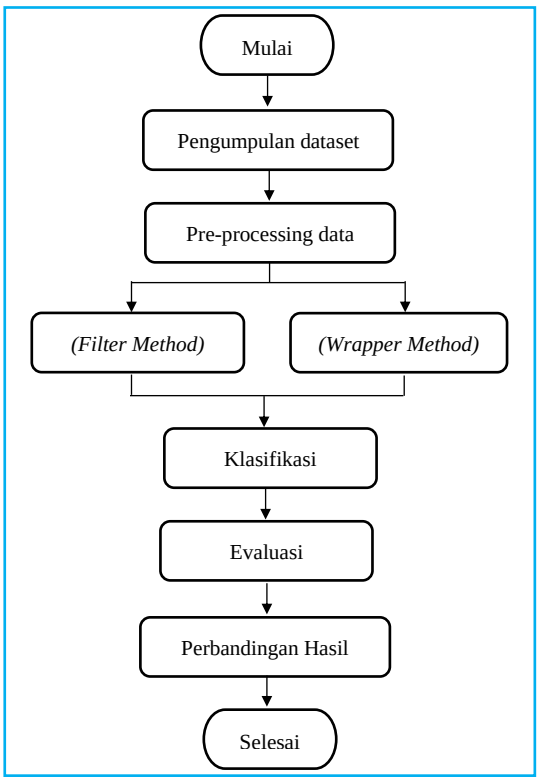
Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada peningkatan akurasi tanpa mempertimbangkan isu ketidakseimbangan kelas yang umum terjadi pada dataset medis, serta masih terbatas dalam penggunaan metrik evaluasi seperti ROC-AUC dan recall kelas minoritas. Padahal, dalam konteks klinis, kesalahan *false negative* pada kelas minoritas (pasien

 ISSN:2302-8149
e-ISSN:2540-9719

berisiko) jauh lebih kritis dibandingkan kesalahan pada kelas mayoritas. Selain itu, studi yang secara langsung membandingkan *filter method* dan *wrapper method* pada klasifikasi penyakit jantung berbasis k-NN dalam kondisi *data imbalanced* masih jarang dilakukan. Oleh karena itu, penelitian ini mengisi celah tersebut dengan mengevaluasi kedua metode seleksi fitur secara komprehensif dalam konteks prediksi risiko penyakit jantung.

## 3 Metode Penelitian

Metodologi penelitian ini difokuskan pada perbandingan dua pendekatan seleksi fitur, yaitu ***filter method*** dan ***wrapper method*** dalam klasifikasi risiko penyakit jantung menggunakan algoritma *k-Nearest Neighbors* (*k*-NN). *Filter method* akan menerapkan perhitungan *correlation* sedangkan *wrapper method* akan menerapkan *Sequential Forward Selection*. Tahapan penelitian meliputi pengumpulan dan *preprocessing* data, pembagian dataset, pelatihan model, serta evaluasi performa untuk menilai efektivitas masing-masing metode seleksi fitur. Alur lengkap tahapan penelitian ini dapat dilihat pada Gambar 1.




**Gambar 1 Tahapan penelitian**

### 3.1 Dataset

Dalam penelitian ini, data diperoleh dari *platform Kaggle*. Dataset ini merupakan dataset publik yang tersedia secara terbuka dan digunakan sebagai data sekunder untuk eksperimen prediksi risiko penyakit jantung. Dataset ini berisi 10.000 entri pasien dengan fitur-fitur yang berkaitan dengan status penyakit jantung dan berbagai faktor risiko, sebagaimana ditunjukkan pada Tabel 1.

**Tabel 1 Deskripsi fitur**

| Fitur | Deskripsi | Tipe Data |
|---|---|---|
| Age | Usia pasien (dalam tahun) | Integer |
| Gender | Jenis kelamin pasien (Male/Female) | Kategorikal |
| Blood_Pressure | Tekanan darah sistolik (mmHg) | Integer |
| Cholesterol_Level | Kadar kolesterol total (mg/dL) | Integer |
| Exercise_Habits | Frekuensi olahraga (High/Medium/Low) | Kategorikal |
| Smoking | Kebiasaan merokok (Yes/No) | Boolean/Kategorikal |
| Family_Heart_Disease | Riwayat penyakit jantung dalam keluarga (Yes/No) | Boolean/Kategorikal |

 ISSN:2302-8149
e-ISSN:2540-9719

| Fitur | Deskripsi | Tipe Data |
|---|---|---|
| Diabetes | Riwayat diabetes (Yes/No) | Boolean/Kategorikal |
| BMI | Indeks massa tubuh (kg/m²) | Float |
| High_Blood_Pressure | Diagnosa hipertensi (Yes/No) | Boolean/Kategorikal |
| Low_HDL_Cholesterol | Kolesterol HDL rendah (Yes/No) | Boolean/Kategorikal |
| High_LDL_Cholesterol | Kolesterol LDL tinggi (Yes/No) | Boolean/Kategorikal |
| Alcohol_Consumption | Tingkat konsumsi alkohol (Low/Medium/High) | Kategorikal |
| Stress_Level | Tingkat stres (Low/Medium/High) | Kategorikal |
| Sleep_Hours | Jam tidur rata-rata per hari | Float |
| Sugar_Consumption | Tingkat konsumsi gula (Low/Medium/High) | Kategorikal |
| Triglyceride_Level | Kadar trigliserida (mg/dL) | Integer |
| Fasting_Blood_Sugar | Gula darah puasa (mg/dL) | Integer |
| CRP_Level | C-reactive protein (mg/L) | Float |
| Homocysteine_Level | Kadar homosistein dalam darah (µmol/L) | Float |
| Heart_Disease_Status | Status penyakit jantung (Yes/No) | Boolean/Kategorikal |

### 3.2 Preprocessing

*Preprocessing* dilakukan untuk menyiapkan dataset penyakit jantung yang terdiri dari 10.000 data poin dan 21 fitur yang merepresentasikan berbagai indikator kesehatan, gaya hidup, dan kondisi biologis. Tujuan utama *preprocessing* adalah mengatasi ketidakteraturan data dan meningkatkan performa model prediksi [11]. Langkah-langkah *preprocessing* adalah sebagai berikut:

a. Pemeriksaan dan Penanganan Data Hilang

Dataset dianalisis untuk mendeteksi keberadaan nilai kosong maupun *outlier* pada setiap fitur. Penanganan nilai hilang dilakukan secara berbeda sesuai dengan jenis fitur. Untuk variabel numerik, digunakan *mean imputation* agar distribusi data tetap terjaga serta tidak terdistorsi oleh nilai ekstrem. Sementara itu, pada variabel kategorikal seperti *Smoking* atau *Diabetes*, dilakukan *mode imputation*, yaitu menggantikan nilai kosong dengan kategori yang paling sering muncul. Strategi ini dipilih karena sederhana, efektif, dan mampu mempertahankan karakteristik asli data tanpa mengurangi kualitas informasi [12].

b. Transformasi Data

Tahap transformasi data dilakukan untuk menyeragamkan format data yang dilanjutkan dengan tahap normalisasi data untuk menyamakan skala nilai pada setiap fitur sebelum digunakan dalam pemodelan. Hal ini diperlukan agar variabel dengan rentang besar tidak mendominasi perhitungan, terutama pada algoritma berbasis jarak seperti *k-Nearest Neighbors* (*k*-NN).

Pada penelitian ini, fitur numerik dinormalisasi menggunakan *Min-Max Normalization*, yaitu mengubah nilai ke dalam rentang [0,1]. Dengan cara ini, perbedaan skala antarvariabel dapat diseimbangkan tanpa mengubah distribusi relatif data. Rumus *min-max* ditunjukan dalam persamaan (1).

$$x^I = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Selain itu, variabel kategorikal seperti *Gender*, *Smoking*, dan *Diabetes* dikonversi ke bentuk numerik melalui *one-hot encoding*, sehingga setiap kategori direpresentasikan dalam variabel *dummy*. Proses transformasi ini diharapkan dapat meningkatkan kualitas input, mempercepat konvergensi model, serta menghasilkan prediksi yang lebih akurat.

c. Pembagian Dataset

Data dibagi menjadi *training set* (80%) dan *testing set* (20%) menggunakan *stratified sampling* untuk memastikan distribusi kelas target tetap konsisten di kedua subset [13]. Pendekatan *stratified sampling* juga berfungsi sebagai strategi mitigasi ketidakseimbangan data, sehingga proporsi kelas mayoritas dan minoritas tetap terjaga pada data latih dan uji. Pembagian data dilakukan menggunakan skema *holdout validation* dengan parameter *random_state* = 42 agar eksperimen bersifat *reproducible* dan hasil evaluasi dapat direplikasi. Data ini selanjutnya digunakan untuk membangun dan mengevaluasi model prediksi *k*-NN.

 ISSN:2302-8149
e-ISSN:2540-9719

### 3.3 Fitur Seleksi (Feature Selection)

Seleksi fitur merupakan tahap penting dalam pemodelan *machine learning* karena membantu mengurangi dimensi data, meningkatkan performa, serta mengurangi risiko *overfitting* [14]. Secara umum, metode seleksi fitur dapat dikelompokkan menjadi dua kategori besar, yaitu *filter method* dan *wrapper method*. *Filter method* melakukan pemilihan fitur berdasarkan karakteristik statistik masing-masing variabel secara independen terhadap target menggunakan korelasi dan *Chi-Square*. Kelebihan metode ini adalah prosesnya yang cepat dan efisien, meskipun terkadang mengabaikan interaksi antar fitur.

Sebaliknya, *wrapper method* menggunakan algoritma pembelajaran mesin secara langsung dalam proses pemilihan fitur. Fitur dipilih dengan mempertimbangkan dampaknya terhadap kinerja model, sehingga interaksi antar fitur dapat diperhitungkan. Salah satu pendekatan *wrapper method* yang banyak digunakan adalah *Sequential Forward Selection* (SFS). Metode ini bekerja secara iteratif dengan memulai dari himpunan fitur kosong, kemudian menambahkan fitur satu per satu berdasarkan kontribusi terbaiknya terhadap peningkatan performa model [15]. Proses berlanjut hingga tidak ada lagi peningkatan signifikan atau jumlah fitur optimal tercapai. Dengan demikian, SFS tidak hanya menyederhanakan model, tetapi juga meningkatkan generalisasi dan mengurangi risiko *overfitting*.

*a. Correlation*

Korelasi merupakan ukuran statistik yang menunjukkan derajat hubungan linear antara dua variabel. Dalam konteks seleksi fitur, korelasi digunakan untuk mengetahui seberapa kuat fitur numerik berhubungan dengan target biner [16]. Salah satu bentuk korelasi yang digunakan adalah *Point-Biserial Correlation*, yang merupakan turunan dari *Pearson Correlation* khusus untuk variabel numerik terhadap variabel dikotomi (biner). Nilai korelasi *r* berkisar dari -1 hingga +1, di mana nilai positif menunjukkan hubungan searah, nilai negatif menunjukkan hubungan berlawanan, dan nilai mendekati 0 menunjukkan tidak adanya hubungan linear yang kuat.

Rumus *Pearson's correlation coefficient* ditunjukan dalam persamaan (2).

$$r = \frac{\sum(x_i - \bar{x})(y_i - \hat{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \hat{y})^2}} \quad (2)$$

Keterangan:

$x_i, y_i$= nilai data ke-i

$\bar{x}$, ŷ = rata-rata dari *x* dan *y*

Pada seleksi fitur, fitur dengan nilai korelasi lebih tinggi (positif maupun negatif) dianggap lebih relevan untuk prediksi target. Namun, karena korelasi hanya menangkap hubungan linear, ia kurang efektif jika hubungan antar variabel bersifat non-linear.

*b. Chi-Square Test*

Uji *Chi-Square* ($\chi 2$) merupakan metode statistik yang digunakan untuk mengukur asosiasi antara dua variabel kategorikal. Dalam seleksi fitur, *Chi-Square* digunakan untuk mengevaluasi apakah distribusi kategori suatu fitur independen terhadap distribusi target kategorikal [17]. Jika *p-value* dari hasil uji *Chi-Square* kecil (biasanya < 0.05), maka dapat disimpulkan bahwa terdapat hubungan signifikan antara fitur dan target [18].

Rumus statistik *Chi-Square* ditunjukan dalam persamaan (3).

$$x^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (3)$$

$O_{ij}$ = frekuensi observasi pada sel ke-i,j

$E_{ij}$ = frekuensi harapan pada sel ke-i,j

$$E_{ij} = \frac{(total\ baris\ i)(total\ kolom\ j)}{N}$$

di mana $O_{ij}$ adalah frekuensi observasi pada sel ke- *i,j* dan $E_{ij}$ adalah frekuensi harapan pada sel yang sama. Nilai *Chi-Square* besar menunjukkan perbedaan signifikan antara distribusi observasi dan distribusi harapan, yang berarti fitur tersebut berhubungan dengan target. Metode ini hanya cocok untuk data kategorikal, karena mengandalkan tabel kontingensi.

c. *Sequential Forward Selection* (SFS)

*Sequential Forward Selection* (SFS) adalah salah satu *metode feature selection* berbasis *wrapper* yang digunakan untuk memilih subset fitur terbaik dari sekumpulan fitur awal. Prinsip utama SFS adalah memulai dari subset kosong kemudian secara bertahap menambahkan fitur satu per satu yang paling meningkatkan performa model [19]. Proses ini dilakukan dengan strategi *greedy* (rakus), yaitu

 ISSN:2302-8149
e-ISSN:2540-9719

pada setiap langkah hanya dipilih fitur yang memberikan perbaikan terbaik terhadap fungsi evaluasi saat itu. Dengan demikian, SFS tidak menjamin menemukan solusi global optimum, tetapi umumnya menghasilkan kombinasi fitur yang cukup baik dengan ukuran subset lebih kecil sehingga model menjadi lebih sederhana dan efisien.

Secara konseptual, SFS bekerja dengan membandingkan performa model menggunakan berbagai kandidat fitur, lalu memilih fitur yang memberikan peningkatan terbesar ketika ditambahkan ke subset yang sudah ada. Evaluasi ini dilakukan menggunakan fungsi objektif tertentu, misalnya akurasi, *F1-score*, atau AUC. Misalnya jika menggunakan akurasi, maka rumus evaluasi yang digunakan ditunjukan pada persamaan (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Pada setiap iterasi, fitur terbaik ditentukan dengan persamaan (5).

$$f_j^* = arg\,max\ J(S \cup \{f_j\}) \quad (5)$$

dengan $f$ adalah himpunan semua fitur, $S$ subset fitur terpilih, dan $J(\cdot)$fungsi evaluasi. Proses berlanjut hingga jumlah fitur yang diinginkan tercapai atau ketika tidak ada lagi peningkatan signifikan pada performa model.

Metode SFS memiliki sejumlah kelebihan, seperti kesederhanaan, fleksibilitas karena bisa digunakan dengan berbagai algoritma klasifikasi maupun regresi, serta kemampuannya menghasilkan subset fitur yang cukup representatif [20]. Namun, SFS juga memiliki keterbatasan, yakni komputasi yang relatif mahal pada dataset besar serta risiko terjebak pada solusi lokal karena sifatnya yang *greedy*.

Dalam praktiknya, SFS banyak digunakan pada berbagai kasus di bidang kesehatan, bioinformatika, pengenalan pola, hingga industri. Contohnya dalam penelitian medis, SFS dapat membantu memilih variabel klinis paling relevan untuk memprediksi risiko serangan jantung, seperti tekanan darah, kadar kolesterol, indeks massa tubuh, dan kebiasaan olahraga. Dalam bidang pengolahan citra, SFS dapat digunakan untuk memilih subset piksel atau fitur tekstur yang optimal dalam sistem pengenalan wajah. Sedangkan di industri manufaktur, SFS bermanfaat untuk memilih sensor paling relevan dalam sistem *predictive maintenance*. Dengan demikian, SFS merupakan metodologi yang penting dalam pengembangan model *machine learning* yang lebih efisien dan akurat.

### 3.4 Klasifikasi k-Nearest Neighbors

Metode *k-Nearest Neighbors* (k-NN) merupakan algoritma klasifikasi non-parametrik berbasis *instance-based learning*. Algoritma ini bekerja dengan menghitung jarak antar data dan menentukan kelas suatu data baru berdasarkan mayoritas kelas dari *k* tetangga terdekat. Jarak yang digunakan dalam penelitian ini adalah *Euclidean Distance*, karena seluruh fitur yang digunakan dalam pemodelan telah direpresentasikan dalam bentuk numerik dan dinormalisasi menggunakan *min-max scaling*, sehingga berada pada skala yang sebanding. Oleh karena itu, *Euclidean Distance* menjadi metode jarak yang sesuai.

Dalam penelitian ini, k-NN digunakan sebagai model klasifikasi risiko serangan jantung, dengan fokus utama pada tahap seleksi fitur. Seleksi fitur dilakukan agar model k-NN hanya menggunakan fitur yang paling relevan, sehingga dapat mengurangi redundansi, meningkatkan efisiensi komputasi, serta memperbaiki performa prediksi. Fitur klinis seperti tekanan darah, kadar gula darah puasa, indeks massa tubuh (BMI), dan faktor risiko lainnya dievaluasi menggunakan metode korelasi untuk data numerik serta *Chi-square test* untuk data kategorikal.

Tahapan penerapan k-NN dalam penelitian ini meliputi:

*Preprocessing* Data: Pembersihan nilai hilang, normalisasi fitur numerik dengan *min-max scaling*, serta *encoding* variabel kategorikal.

Seleksi Fitur: Menggunakan korelasi dan uji *Chi-square* untuk memilih fitur yang berkontribusi signifikan terhadap status penyakit jantung.

Pembangunan Model k-NN: Model k-NN pada penelitian ini dibangun dengan nilai parameter $k$ = 5, yang ditetapkan sebagai parameter tetap mengacu pada praktik umum dan hasil penelitian sebelumnya yang menunjukkan bahwa nilai $k$ = 5 mampu memberikan performa klasifikasi yang lebih

 ISSN:2302-8149
e-ISSN:2540-9719

stabil dibandingkan nilai *k* yang lebih kecil. Nilai *k* tersebut digunakan secara konsisten pada seluruh proses pelatihan dan evaluasi tanpa dilakukan *hyperparameter tuning* seperti *grid search* karena fokus utama penelitian ini adalah menganalisis pengaruh seleksi fitur *filter method* dan *wrapper method* terhadap performa model, bukan mengoptimasi parameter *classifier*. Dengan penggunaan parameter tetap, dampak seleksi fitur dapat diamati secara lebih objektif dan terkontrol, dengan fungsi jarak *Euclidean Distance*, yang dirumuskan sebagaimana persamaan 6 berikut:

$$D = \sqrt{\sum_{i=1}^{n}(x_1 - y_1)^2} \quad (6)$$

Keterangan:
*x* = contoh data
*y* = data uji
*i* = variabel data
*n* = jumlah data latih

Selain itu, model k-NN ini menggunakan parameter *weights* = uniform untuk memberikan bobot yang sama pada setiap tetangga terdekat, serta *algorithm* = auto agar pemilihan metode pencarian tetangga dilakukan secara otomatis oleh sistem.

### 3.5 Evaluasi

Dataset memiliki distribusi kelas yang tidak seimbang, sehingga evaluasi model tidak hanya mengandalkan akurasi. Untuk itu, digunakan ROC-AUC dan recall kelas minoritas sebagai metrik evaluasi. Selain mitigasi *imbalance* melalui *stratified sampling* pada pembagian data, ROC-AUC menilai performa model pada seluruh rentang *threshold* dan recall minoritas mengevaluasi kemampuan model mendeteksi pasien positif yang jarang. Kombinasi kedua metrik ini memastikan evaluasi yang representatif terhadap kedua kelas.

*a. Confusion Matrix*

*Confusion Matrix* adalah alat evaluasi kinerja model klasifikasi dalam *machine learning* yang menunjukkan jumlah prediksi yang benar dan salah yang dibuat oleh model dibandingkan dengan label aktual. Matriks ini terdiri dari empat komponen utama yang digunakan untuk menghitung metrik evaluasi lainnya, seperti akurasi, presisi, *recall*, dan *F1-score*. Struktur *confusion matrix* untuk masalah klasifikasi biner (2 kelas) ditunjukkan pada Tabel 2.

**Tabel 2 *Confusion matrix***

| | **Prediksi Positif** | **Prediksi Negatif** |
|---|---|---|
| **Aktual Positif** | True Positif (TP) | False Negatif (FN) |
| **Aktual Negatif** | False Positif (FP) | True Negatif (TN) |

Penjelasan:
TP (*True Positive*) : Model memprediksi positif, dan data sebenarnya memang positif.
TN (*True Negative*) : Model memprediksi negatif, dan data sebenarnya memang negatif.
FP (*False Positive*) : Model memprediksi positif, padahal data sebenarnya negatif.
FN (*False Negative*) : Model memprediksi negatif, padahal data sebenarnya positif.

b. ROC-AUC

*Receiver Operating Characteristic – Area Under the Curve* (ROC-AUC) merupakan salah satu metrik evaluasi yang banyak digunakan pada klasifikasi biner, terutama ketika data memiliki distribusi kelas yang tidak seimbang.

ROC adalah kurva yang menggambarkan hubungan antara *True Positive Rate* (*TPR/Recall*) dengan *False Positive Rate* (FPR) pada berbagai nilai ambang batas (*threshold*) probabilitas. Rumus perhitungan kedua metrik tersebut ditunjukkan pada persamaan (7) dan persamaan (8).

- TPR (*Recall*) = TP / (TP + FN) (7)
- FPR = FP / (FP + TN) (8)

Dengan mengubah nilai ambang batas, diperoleh serangkaian titik koordinat (TPR, FPR) yang kemudian membentuk kurva ROC.

Luas di bawah kurva ROC disebut AUC (*Area Under the Curve*), yang menjadi indikator seberapa baik model mampu membedakan kelas positif dan negatif. Interpretasi nilai ROC-AUC adalah sebagai berikut:

 ISSN:2302-8149
e-ISSN:2540-9719

- AUC = 1,0 → model sempurna dalam klasifikasi
- AUC = 0,5 → model tidak lebih baik dari tebakan acak
- 0,5 < AUC < 1,0 → semakin besar nilai AUC, semakin baik kemampuan model dalam memisahkan kelas

Secara umum, kategori interpretasi nilai AUC dapat dibagi:

- 0,5 → tidak ada kemampuan diskriminasi *(random)*
- 0,6–0,7 → cukup (*fair*)
- 0,7–0,8 → baik (*good*)
- 0,8–0,9 → sangat baik (*very good*)
- 0,9 → hampir sempurna *(excellent*)

Dengan mengandalkan ROC-AUC dan recall kelas minoritas, evaluasi model menjadi lebih representatif dan sensitif terhadap ketidakseimbangan data. ROC-AUC tidak bergantung pada satu titik ambang batas tertentu, sehingga memberikan gambaran performa model yang menyeluruh pada seluruh rentang *threshold*. Pendekatan ini juga memastikan kemampuan model dalam mendeteksi kelas minoritas tetap terukur dengan baik, menjadikan kombinasi metrik ini pilihan tepat dalam penelitian medis dengan distribusi kelas yang tidak seimbang.

## 4 Hasil dan Pembahasan

### 4.1 Pengumpulan Data

Dalam penelitian ini, data diperoleh dari *platform Kaggle*. Dataset ini berisi 10.000 entri pasien dengan fitur-fitur yang berkaitan dengan status penyakit jantung dan berbagai faktor risiko, seperti tekanan darah, kadar kolesterol, kebiasaan merokok, konsumsi alkohol, dan riwayat keluarga.
Dataset semula tersimpan dalam format satu kolom yang memuat data dalam bentuk teks terpisah titik koma (;). Oleh karena itu, preprocessing awal dilakukan dengan cara split kolom, pembersihan baris header palsu, serta penyesuaian tipe data untuk mendukung proses klasifikasi.

### 4.2 Preprocessing

a. Pemeriksaan dan Penanganan Data Hilang

Hasil analisis pada Tabel 3 menunjukkan adanya jumlah data hilang pada setiap fitur. Secara umum, *missing value* tersebar pada hampir seluruh variabel input, meskipun jumlahnya relatif kecil dibandingkan total data. Seluruh data telah memiliki label pada variabel target *Heart Disease Status*, menandakan bahwa tidak terdapat nilai yang hilang (*missing value*) dan konsistensi dataset terjaga. Keberadaan data hilang ini berpotensi menurunkan kualitas analisis apabila tidak ditangani dengan tepat. Oleh karena itu, strategi imputasi diperlukan. Untuk variabel numerik digunakan *mean imputation* agar distribusi data tetap stabil, sedangkan variabel kategorikal ditangani dengan *mode imputation*.

**Tabel 3 Jumlah *missing value***

| Fitur | Jumlah *Missing Value* |
|---|---|
| Age | 29 |
| Gender | 19 |
| Blood_Pressure | 19 |
| Cholesterol_Level | 30 |
| Exercise_Habits | 25 |
| Smoking | 25 |
| Family_Heart_Disease | 21 |
| Diabetes | 30 |
| BMI | 22 |
| High_Blood_Pressure | 26 |
| Low_HDL_Cholesterol | 25 |
| High_LDL_Cholesterol | 26 |
| Alcohol_Consumption | 32 |

 ISSN:2302-8149
e-ISSN:2540-9719

| Fitur | Jumlah *Missing Value* |
|---|---|
| Stress_Level | 22 |
| Sleep_Hours | 25 |
| Sugar_Consumption | 30 |
| Triglyceride_Level | 26 |
| Fasting_Blood_Sugar | 22 |
| CRP_Level | 26 |
| Homocysteine_Level | 20 |
| Heart_Disease_Status | 0 |

Berdasarkan Tabel 3, dapat disimpulkan bahwa sebagian besar fitur memiliki *missing value* dalam jumlah terbatas, sehingga diperlukan penanganan untuk menjaga kualitas dataset sebelum tahap pemodelan. Kondisi ini menunjukkan bahwa proses imputasi menjadi langkah penting untuk mencegah hilangnya informasi serta mengurangi potensi bias dalam proses klasifikasi.

Untuk mengatasi permasalahan *missing value* dilakukan proses imputasi dengan pendekatan yang berbeda pada data numerik dan kategorikal. Pertama, seluruh kolom numerik dikonversi ke tipe *float* agar dapat dihitung nilai *mean*-nya secara tepat. Selanjutnya, nilai hilang pada kolom numerik diisi menggunakan rata-rata (*mean imputation*). Pemilihan metode *mean* dilakukan karena dianggap mampu menjaga pola distribusi data, sehingga representasi umum dari fitur tersebut tetap terjaga.

Sementara itu, untuk fitur kategorikal, seperti *Smoking*, *Diabetes*, atau *Exercise Habits*, digunakan metode imputasi dengan modus (*mode imputation*). Artinya, nilai yang hilang digantikan oleh kategori yang paling sering muncul pada fitur tersebut. Metode ini dipilih karena kategori dengan frekuensi tertinggi dinilai paling merepresentasikan kecenderungan data asli.

Dengan penerapan kombinasi *mean imputation* untuk data numerik dan *mode imputation* untuk data kategorikal, kualitas dataset meningkat dan risiko bias akibat data hilang dapat diminimalkan. Hasil ini memastikan dataset siap digunakan untuk tahap analisis berikutnya, seperti seleksi fitur dan pembangunan model klasifikasi.

### 4.3 Fitur Seleksi (*Feature Selection*)

*a. Filter Method*

Hasil seleksi fitur dengan *filter method* pada data numerik yang menggunakan korelasi menunjukkan bahwa sebagian besar fitur memiliki nilai *r* yang relatif kecil terhadap status penyakit jantung. Hal ini mengindikasikan bahwa hubungan linear masing-masing variabel numerik secara individu tidak terlalu kuat dalam menjelaskan risiko penyakit jantung.

Dalam penelitian ini, ditetapkan batas *threshold* $|r| \geq 0{,}01$ berdasarkan heuristik eksploratif untuk mempertahankan fitur yang dianggap masih memiliki kontribusi meskipun kecil [21]. Dengan ambang ini, hanya fitur seperti *BMI*, *Homocysteine Level* dan *Blood Pressure* yang dipertahankan, sementara variabel lain dengan nilai di bawah *threshold* dieliminasi karena dianggap tidak memberikan informasi signifikan.

Secara keseluruhan, hasil pada Tabel 4 menegaskan bahwa pendekatan korelasi memiliki keterbatasan karena hanya mengevaluasi kekuatan hubungan satu variabel terhadap target tanpa mempertimbangkan kombinasi faktor klinis yang lebih kompleks.

**Tabel 4 Nilai r**

| Fitur | *Nilai r* |
|---|---|
| **BMI** | **0.019682** |
| **Homocysteine_Level** | **0.008296** |
| Triglyceride_Level | 0.002914 |
| Cholesterol_Level | 0.002706 |
| Fasting_Blood_Sugar | -0.002244 |
| Sleep_Hours | -0.003816 |
| CRP_Level | -0.006009 |

 ISSN:2302-8149
e-ISSN:2540-9719

| | |
|---|---|
| Age | -0.009239 |
| **Blood_Pressure** | **-0.013877** |

Pada fitur kategorikal, hasil uji signifikansi menggunakan *p-value* memperlihatkan bahwa sebagian besar variabel tidak signifikan terhadap status penyakit jantung, karena memiliki *p-value* di atas 0,05 sebagaimana Tabel 5. Namun, *Stress Level* terbukti signifikan, sehingga dapat dianggap sebagai salah satu faktor penting yang berhubungan dengan risiko penyakit jantung. Temuan ini menunjukkan bahwa aspek psikologis dan gaya hidup dapat memberikan kontribusi yang lebih bermakna dibandingkan beberapa variabel klinis yang tidak signifikan secara statistik.

**Tabel 5 *p-value***

| **Fitur** | ***p-value*** |
|---|---|
| Gender | 0.0901 |
| Exercise_Habits | 0.7863 |
| Smoking | 0.8064 |
| Family_Heart_Disease | 0.4684 |
| Diabetes | 0.8064 |
| High_Blood_Pressure | 0.8454 |
| Low_HDL_Cholesterol | 0.5653 |
| High_LDL_Cholesterol | 0.4266 |
| **Stress_Level** | **0.0240** |

Hasil seleksi fitur dengan *filter method* menunjukkan bahwa hanya beberapa fitur yang benar-benar relevan untuk klasifikasi. Pada variabel numerik, *BMI, Homocysteine Level dan Blood Pressure* terpilih melalui uji korelasi, sedangkan pada variabel kategorikal, *Stress Level* terpilih melalui uji *Chi-Square*. Dengan demikian, *filter method* dalam penelitian ini dilakukan melalui dua pendekatan, yaitu korelasi untuk variabel numerik dan *Chi-Square* untuk variabel kategorikal. Subset akhir kemudian diperoleh dengan menggabungkan fitur-fitur yang lolos dari kedua pendekatan tersebut, sehingga diharapkan dapat mengurangi *noise* dan meningkatkan performa model.

b. *Wrapper Method* (*Sequential Forward Selection*)

Penerapan metode *Sequential Forward Selection* (SFS) pada dataset menghasilkan pemilihan 11 (sebelas) fitur terbaik yang dianggap paling relevan terhadap klasifikasi risiko penyakit jantung. Berbeda dengan *filter method*, SFS mengevaluasi fitur secara bertahap berdasarkan kontribusi kombinasi variabel terhadap performa model, sehingga interaksi antar fitur dapat diperhitungkan.

Berdasarkan hasil iterasi pada Tabel 6, penambahan fitur secara bertahap menghasilkan performa model yang relatif stabil dengan peningkatan akurasi dibandingkan subset fitur hasil *filter method*. Hal ini menunjukkan bahwa pendekatan *wrapper method* lebih efektif dalam memilih kombinasi faktor risiko yang saling melengkapi.

**Tabel 6 Iterasi**

| **Iterasi ke-** | **Akurasi** |
|---|---|
| 1. | 0.8001 |
| 2. | 0.8001 |
| 3. | 0.8003 |
| 4. | 0.8001 |
| 5. | 0.8001 |
| 6. | 0.8001 |
| 7. | 0.7999 |
| 8. | 0.8004 |
| 9. | 0.8006 |
| 10. | 0.8000 |
| **Rata-rata** | **0.8001** |

Temuan ini mengindikasikan bahwa SFS efektif dalam memilih fitur dengan mempertimbangkan interaksi antar variabel, bukan hanya hubungan satu arah dengan target seperti pada *filter method*.

 ISSN:2302-8149
e-ISSN:2540-9719

Misalnya, kombinasi antara *BMI*, *Homocysteine Level*, dan *High LDL Cholesterol* memberikan kontribusi penting dalam mendeteksi risiko penyakit jantung, yang sesuai dengan literatur medis terkait faktor obesitas, metabolisme, serta gangguan lipid. Selain itu, keterpilihan fitur *Stress Level* dan *Sleep Hours* menunjukkan bahwa faktor gaya hidup juga memiliki peran yang tidak dapat diabaikan.

Dengan demikian, dapat disimpulkan bahwa SFS tidak hanya membantu menyederhanakan jumlah fitur dari keseluruhan dataset, tetapi juga meningkatkan akurasi model. Keunggulan utama metode ini terletak pada kemampuannya mengevaluasi kombinasi fitur yang saling melengkapi, sehingga menghasilkan model klasifikasi yang lebih stabil, efisien, dan memiliki daya prediksi lebih baik dibandingkan pendekatan *filter method*.

### 4.4 Klasifikasi K-Nearest Neighbors

Pada tahap awal, model *k-Nearest Neighbors* (*k*-NN) dibangun menggunakan keseluruhan 20 fitur yang tersedia. Langkah ini dimaksudkan sebagai *baseline* untuk mengetahui performa dasar model sebelum dilakukan penyederhanaan fitur. Dalam tahap *preprocessing*, telah diterapkan *stratified sampling* sebagai langkah awal untuk menjaga proporsi kelas mayoritas dan minoritas pada data pelatihan dan pengujian. Namun demikian, dataset masih memiliki karakteristik tidak seimbang (*imbalanced data*), sehingga berpotensi menimbulkan bias prediksi terhadap kelas mayoritas. Hasil pengujian menunjukkan bahwa model baseline memiliki kecenderungan lebih akurat pada kelas mayoritas (pasien sehat), sementara kemampuan mendeteksi kelas minoritas (pasien sakit) masih sangat rendah. Kondisi ini mencerminkan keterbatasan algoritma berbasis jarak seperti k-NN, di mana keputusan klasifikasi sangat dipengaruhi oleh dominasi kelas mayoritas dalam lingkungan tetangga terdekat. Hal ini menunjukkan bahwa penerapan *stratified sampling* saja belum cukup untuk sepenuhnya mengatasi dampak ketidakseimbangan kelas. Selain itu, meskipun performa global model dapat terlihat cukup baik, recall kelas minoritas tetap rendah karena keputusan klasifikasi masih didasarkan pada *threshold* standar, sehingga banyak kasus pasien berisiko masih terklasifikasikan sebagai kelas mayoritas (*false negative*).

Untuk mengurangi dampak tersebut, dilakukan seleksi fitur dengan *filter method*. Proses ini melalui dua tahap: analisis korelasi untuk fitur numerik dan uji *Chi-Square* untuk fitur kategorikal. Dari hasil analisis, hanya empat fitur yang signifikan, yaitu *BMI*, *Homocysteine Level*, *Blood Pressure*, dan *Stress Level*. Penggunaan subset fitur ini hanya memberikan peningkatan performa yang terbatas, terutama pada kemampuan model mengenali kelas minoritas. Hal ini sejalan dengan karakteristik *filter method* yang mengevaluasi fitur secara individual tanpa mempertimbangkan interaksi antar faktor risiko penyakit jantung [6].

Tahap berikutnya adalah penerapan *wrapper method* menggunakan teknik *Sequential Forward Selection* (SFS). Metode ini mengevaluasi kombinasi fitur secara iteratif, menambahkan satu demi satu fitur yang paling meningkatkan performa model. Proses ini menghasilkan 11 fitur terbaik yang mencakup faktor klinis maupun gaya hidup. Keunggulan SFS dibandingkan *filter method* terletak pada kemampuannya mempertimbangkan dependensi dan interaksi antar fitur, yang sangat relevan pada kasus penyakit jantung yang bersifat multifaktorial [6][19]. Model *k*-NN dengan subset fitur hasil SFS menunjukkan akurasi rata-rata 80,00% dan nilai ROC-AUC sebesar 0,657. Peningkatan ini mengindikasikan bahwa pemilihan kombinasi fitur yang optimal mampu mengurangi dominasi kelas mayoritas dan meningkatkan kemampuan model dalam mengenali pasien berisiko, meskipun teknik penyeimbangan data belum diterapkan secara eksplisit. Namun, peningkatan ROC-AUC ini lebih merefleksikan perbaikan kemampuan model dalam melakukan ranking dan diskriminasi antar kelas secara keseluruhan, sedangkan recall minoritas masih belum optimal karena ketidakseimbangan kelas tetap memengaruhi prediksi pada ambang keputusan tertentu.

Keterpilihan fitur-fitur tertentu juga memiliki justifikasi medis yang kuat. BMI berkaitan erat dengan obesitas dan risiko kardiovaskular, kadar homosistein dan LDL kolesterol merupakan biomarker yang berhubungan dengan aterosklerosis, sementara faktor gaya hidup seperti stres dan durasi tidur berkontribusi terhadap kondisi kardiometabolik pasien. Temuan ini sejalan dengan penelitian [9] yang menunjukkan bahwa kombinasi faktor klinis dan gaya hidup lebih efektif dalam memprediksi risiko penyakit jantung dibandingkan evaluasi fitur secara terpisah, serta didukung oleh kajian [19] yang menekankan pentingnya metode seleksi fitur yang mampu menangkap interaksi antar variabel pada prediksi risiko penyakit. Selain itu, ulasan komprehensif pada [6] juga melaporkan

 ISSN:2302-8149
e-ISSN:2540-9719

bahwa variabel klinis seperti BMI dan tekanan darah merupakan prediktor yang paling sering digunakan dan relevan dalam prediksi penyakit jantung, sehingga memperkuat validitas subset fitur yang terpilih dalam penelitian ini.

### 4.5 Evaluasi

*a. Confusion Matrix*

Evaluasi kinerja model *k*-NN sebelum dan sesudah seleksi fitur menunjukkan bahwa tahap seleksi fitur memberikan dampak positif terhadap kualitas prediksi, meskipun peningkatannya tidak terlalu besar pada nilai akurasi. Sebelum seleksi fitur, nilai recall kelas positif (pasien sakit) masih sangat rendah, yang menunjukkan tingginya risiko false negative. Dalam konteks medis, kesalahan false negative memiliki implikasi klinis yang serius karena pasien berisiko dapat terlewat dan tidak memperoleh pemeriksaan lanjutan yang diperlukan.

Setelah seleksi fitur dilakukan, performa model pada kelas positif mengalami perbaikan. Meskipun peningkatan akurasi secara keseluruhan relatif kecil, seleksi fitur memberikan kontribusi lebih penting dalam meningkatkan sensitivitas model terhadap pasien berisiko. Hal ini menunjukkan bahwa pendekatan *wrapper method* lebih efektif dalam menekan risiko *false negative*, sehingga lebih sesuai untuk skenario skrining klinis awal, di mana sensitivitas terhadap pasien berisiko lebih diutamakan dibandingkan akurasi keseluruhan. Ringkasan perubahan metrik performa sebelum dan sesudah seleksi fitur ditunjukkan pada Tabel 7. Tabel tersebut menegaskan bahwa *wrapper method* memberikan peningkatan paling signifikan terutama pada metrik kelas minoritas.

**Tabel 7 Evaluasi metrik**

| Metrik | Sebelum Seleksi Fitur | Seleksi Fitur Filter Method | Seleksi Fitur Wrapper | Perubahan Vs Filter | Perubahan Vs Wrapper |
|---|---|---|---|---|---|
| Akurasi | 0.7635 (76,35%) | 0.7680 (76,80%) | 0.8000 (80,00%) | ↑ +0.45% | ↑ +3.65% |
| Precision (kelas 1) | 0.18 | 0.23 | 0.50 | ↑ +0.05 | ↑ +0.32 |
| Recall (kelas 1) | 0.05 | 0.07 | 0.20 | ↑ +0.02 | ↑ +0.15 |
| F1-Score (kelas 1) | 0.08 | 0.11 | 0.29 | ↑ +0.03 | ↑ +0.21 |
| Precision (kelas 0) | 0.8 | 0.8 | 0.83 | stabil | ↑ +0.03 |
| Recall (kelas 0) | 0.94 | 0.94 | 0.95 | stabil | ↑ +0.03 |
| F1-Score (kelas 0) | 0.86 | 0.87 | 0.88 | ↑ +0.01 | ↑ +0.02 |

Dengan demikian, seleksi fitur terbukti bermanfaat dalam mengurangi *noise* dan mempertahankan fitur yang relevan, Namun, hasil ini juga menunjukkan bahwa pendekatan *filter method* memiliki keterbatasan dalam meningkatkan recall kelas minoritas secara signifikan, karena pola risiko penyakit jantung tidak dapat direpresentasikan secara optimal hanya melalui evaluasi fitur secara terpisah.

b. ROC-AUC

Pengukuran performa model klasifikasi juga dilakukan menggunakan metrik *Receiver Operating Characteristic – Area Under Curve* (ROC-AUC). Nilai ROC-AUC tanpa seleksi fitur menunjukkan bahwa kemampuan diskriminasi model masih sangat rendah dan mendekati tebakan acak. Hal ini mengindikasikan bahwa model baseline belum mampu membedakan pasien sehat dan pasien sakit secara efektif. Setelah diterapkan *filter method*, peningkatan ROC-AUC masih sangat terbatas. Hasil ini sejalan dengan temuan pada confusion matrix, di mana *filter method* belum mampu meningkatkan kemampuan model dalam membedakan pasien sehat dan sakit secara efektif.

Perubahan signifikan terlihat setelah menggunakan *wrapper method* (SFS). Nilai ROC-AUC meningkat ke tingkat yang lebih memadai, menunjukkan bahwa model mulai memiliki kemampuan diskriminasi yang cukup untuk kebutuhan skrining risiko penyakit jantung. Secara kuantitatif, nilai ROC-AUC meningkat hingga 0,657, jauh di atas baseline 0,50 yang merepresentasikan tebakan acak. Dalam konteks skrining medis, AUC pada kategori *fair classification* berarti model sudah lebih baik dibandingkan tebakan acak dalam mengidentifikasi individu berisiko untuk pemeriksaan lanjutan. Hal ini penting karena kesalahan *false negative* pada pasien berisiko tinggi dapat menyebabkan keterlambatan penanganan medis.

Meskipun akurasi model dengan *wrapper method* cukup tinggi, perbandingan ROC-AUC ini menegaskan bahwa perbaikan performa terutama datang dari sisi kemampuan model membedakan

 ISSN:2302-8149
e-ISSN:2540-9719

kelas positif dan negatif, bukan sekadar peningkatan prediksi pada kelas mayoritas. Keunggulan utama *wrapper method* terletak pada kemampuannya mengevaluasi interaksi antar fitur melalui proses *Sequential Forward Selection*, sehingga kombinasi fitur klinis dan gaya hidup dapat berkontribusi secara bersama-sama dalam mendeteksi kelas minoritas. Namun, recall kelas minoritas masih relatif rendah karena distribusi kelas yang tidak seimbang menyebabkan model tetap cenderung bias terhadap kelas mayoritas. Meskipun ROC-AUC meningkat yang menunjukkan kemampuan diskriminasi model secara keseluruhan membaik, peningkatan tersebut belum sepenuhnya diikuti oleh sensitivitas terhadap kelas minoritas akibat jumlah sampel minoritas yang terbatas dan threshold klasifikasi standar. Perbandingan nilai ROC-AUC antar pendekatan ditampilkan pada Tabel 8, yang menunjukkan bahwa *wrapper method* memberikan peningkatan kemampuan diskriminasi paling besar dibandingkan *baseline* maupun *filter method*.

**Tabel 8 Perbandingan ROC-AUC**

| Metrik | ROC-AUC | Perubahan vs baseline |
|---|---|---|
| Tanpa Seleksi | 0.5054 | - |
| Filter Method | 0.5111 | ↑ +0.0057 |
| Wrapper (SFS) | 0.6574 | ↑ +0.1520 |

## 5 Kesimpulan

Penelitian ini menunjukkan bahwa pemilihan metode seleksi fitur berpengaruh signifikan terhadap performa klasifikasi risiko penyakit jantung menggunakan algoritma k-Nearest Neighbors (k-NN) pada kondisi dataset medis yang tidak seimbang (*imbalanced*). *Filter method* mampu menyederhanakan jumlah fitur secara efisien, namun peningkatan performanya relatif terbatas. Sebaliknya, *wrapper method* melalui *Sequential Forward Selection* (SFS) menghasilkan subset fitur yang lebih informatif karena mempertimbangkan kombinasi serta interaksi antar fitur, sehingga memberikan peningkatan yang lebih nyata terutama pada nilai ROC-AUC dan kemampuan deteksi kelas minoritas (recall). Kontribusi ilmiah penelitian ini dapat dirangkum sebagai berikut. (1) **Kontribusi teoritis**, penelitian ini membuktikan adanya *trade-off* antara *filter* dan *wrapper* pada data medis *imbalanced*, di mana *filter* lebih sederhana tetapi *wrapper* lebih unggul dalam performa prediksi. (2) **Kontribusi metodologis**, hasil eksperimen menunjukkan bahwa *wrapper method* (SFS) mampu meningkatkan ROC-AUC dan recall kelas minoritas secara lebih signifikan dibandingkan *filter method*, sehingga lebih representatif untuk evaluasi risiko pasien berisiko tinggi. (3) **Kontribusi praktis**, penelitian ini memberikan rekomendasi strategi seleksi fitur yang lebih tepat untuk mendukung pengembangan sistem skrining dan *clinical decision support* pada prediksi penyakit jantung. Meskipun demikian, penelitian ini masih memiliki keterbatasan pada ketidakseimbangan distribusi kelas yang dapat menyebabkan bias terhadap kelas mayoritas dan membatasi kemampuan model dalam mendeteksi seluruh kasus pada kelas minoritas. Oleh karena itu, penelitian lanjutan disarankan untuk menerapkan teknik penyeimbangan data seperti *resampling* atau *SMOTE* sebelum seleksi fitur, serta mengeksplorasi algoritma klasifikasi lain guna meningkatkan performa model dalam konteks pengambilan keputusan medis.

## Referensi

[1] M. Di Cesare *et al.*, "*The Heart of the World*," *Glob. Heart*, Vol. 19, No. 1, 2024, DOI: 10.5334/gh.1288.

[2] M. Fira, L. Goras, and H.-N. Costin, "*Evaluating Sparse Feature Selection Methods: A Theoretical and Empirical Perspective*," *Applied Sciences*, Vol. 15, No. 7, p. 3752, Mar. 2025, DOI: 10.3390/app15073752.

[3] I. G. A. Gunadi and D. O. Rachmawati, "*A Comparative Study on the Impact of Feature Selection and Dataset Resampling on the Performance of the K-Nearest Neighbors (KNN) Classification Algorithm*," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, Vol. 13, No. 2, pp. 419–427, Jul. 2024, DOI: 10.23887/janapati.v13i2.82174.

[4] R. Ishak, "Optimalisasi Seleksi Atribut K-Means menggunakan *Correlation Matrix* pada *Clustering* Penyakit Pasien *Optimization of K-Means Attribute Selection using Correlation*

 ISSN:2302-8149
e-ISSN:2540-9719

*Matrix in Patient Disease Clustering*," *Jambura Journal of Electrical and Electronics Engineering*, Vol. 7, No. 2, Jul. 2025, DOI: 10.37905/jjeee.v7i2.28010.

[5] A. Wantoro, A. Fitria Yulia, D . Yana Ayu, and S. Mustofa, "Evaluasi Kinerja Algoritma *Machine Learning (Ml)* menggunakan Seleksi Fitur pada Klasifikasi Diabetes," *JIP (Jurnal Informatika Polinema)*, Vol. 11, No. 3, May 2025, DOI: 10.33795/jip.v11i3.

[6] F. F. Firdaus, H. A. Nugroho, and I. Soesanti, "*A Review of Feature Selection and Classification Approaches for Heart Disease Prediction*," *International Journal of Information Technology and Electrical Engineering (IJITEE)*, Vol. 4, No. 3, Sep. 2020, DOI: 10.22146/ijitee.59193.

[7] H. Nugroho, G. E. Yuliastuti, and A. Firman, "Klasifikasi Diagnosis *Diabetes Melitus* menggunakan Metode *Naïve Bayes* dengan Seleksi Fitur *Backward Elimination*," *Jurnal Ilmiah NERO*, Vol. 8, No. 2, 2023, DOI: 10.21107/nero.v8i2.21110.

[8] D. Cahya and P. Buani, "Penerapan Algoritma *Naïve Bayes* dengan Seleksi Fitur Algoritma Genetika untuk Prediksi Gagal Jantung," *Jurnal Sains dan Manajemen*, Vol. 9, No. 2, 2021.

[9] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur berbasis *Filter* dan *Wrapper* menggunakan *Naive Bayes* pada Klasifikasi Penyakit Jantung," *Jurnal Teknologi Informasi dan Ilmu Komputer*, Vol. 10, No. 6, pp. 1361–1368, Dec. 2023, DOI: 10.25126/jtiik.2023107467.

[10] Y. Setiawan, "Data Mining berbasis *Nearest Neighbor* dan Seleksi Fitur untuk Deteksi Kanker Payudara," *Jurnal pengembangan IT (JPIT)*, Vol. 8, No. 2, 2023, DOI: 10.30591/jpit.v8i2.4994.

[11] E. N. Wanyonyi and N. W. Masinde, "*The Impact of Data Preprocessing on Machine Learning Model Performance: A Comprehensive Examination*," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, Vol. 11, No. 2, pp. 3814–3827, Apr. 2025, DOI: 10.32628/CSEIT25112854.

[12] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "*An Investigation of the Imputation Techniques for Missing Values in Ordinal Data Enhancing Clustering and Classification Analysis Validity*," *Decision Analytics Journal*, Vol. 9, p. 100341, Dec. 2023, DOI: 10.1016/j.dajour.2023.100341.

[13] Y. Rimal, N. Sharma, S. Paudel, A. Alsadoon, M. P. Koirala, and S. Gill, "*Comparative Analysis of Heart Disease Prediction using Logistic Regression, SVM, KNN, and Random Forest with Cross-Validation for Improved Accuracy*," *SCI. Rep.*, Vol. 15, No. 1, Dec. 2025, DOI: 10.1038/s41598-025-93675-1.

[14] M. S. Pathan, Av. Nag, M. M. Pathan, and S. Dev, "*Analyzing the Impact of Feature Selection on the Accuracy of Heart Disease Prediction*," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.03239

[15] T. Zhao, Y. Zheng, and Z. Wu, "*Feature Selection-based Machine Learning Modeling for Distributed Model Predictive Control of Nonlinear Processes*," *Comput. Chem. Eng.*, Vol. 169, Jan. 2023, DOI: 10.1016/j.compchemeng.2022.108074.

[16] S. Suresh, D. T. Newton, T. H. Everett, G. Lin, and B. S. Duerstock, "*Feature Selection Techniques for a Machine Learning Model to Detect Autonomic Dysreflexia*," *Front. Neuroinform.*, Vol. 16, Aug. 2022, DOI: 10.3389/fninf.2022.901428.

[17] R. R. Sarra, I. I. Gorial, R. R. Manea, A. E. Korial, M. Mohammed, and Y. Ahmed, "*Enhanced Stacked Ensemble-based Heart Disease Prediction with Chi-Square Feature Selection Method*," *Journal of Robotics and Control (JRC)*, Vol. 5, No. 6, pp. 1753–1763, 2024, DOI: 10.18196/jrc.v5i6.23191.

[18] A. I. Neugut and T. Fojo, "*The Statistical Significance Revolution*," *JNCI Cancer Spectr.*, Vol. 8, No. 3, Apr. 2024, DOI: 10.1093/jncics/pkae035.

[19] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, and J. M. O'Sullivan, "*A Review of Feature Selection Methods for Machine Learning-based Disease Risk Prediction*," *Frontiers in Bioinformatics*, Vol. 2, Jun. 2022, DOI: 10.3389/fbinf.2022.927312.

[20] S. Shafiee, L. M. Lied, I. Burud, J. A. Dieseth, M. Alsheikh, and M. Lillemo, "*Sequential Forward Selection and Support Vector Regression in Comparison to LASSO Regression for Spring Wheat Yield Prediction based on UAV Imagery*," *Comput. Electron. Agric.*, Vol. 183, Apr. 2021, DOI: 10.1016/j.compag.2021.106036.

*ISSN:2302-8149*
*e-ISSN:2540-9719*

[21] I. Guyon and A. Elisseeff, "*An Introduction to Variable and Feature Selection*," 2003.