OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

Seleksi Fitur Menggunakan *Sequential Forward Selection* untuk Meningkatkan Kinerja *Decision Tree* dalam Prediksi Batu Empedu

Aries Saifudin[1*], Yulianti[1]

[1]Fakultas Informatika, Program Studi S1 Teknologi Informasi, Direktorat Kampus Jakarta, Universitas Telkom, Jl. Daan Mogot KM.11, RT.1/RW.4, Kedaung Kali Angke, Cengkareng, Kota Jakarta Barat, DKI Jakarta, 11710
Email: [1*]ariessaifudin@telkomuniversity.ac.id, [2]yuliantiy@telkomuniversity.ac.id
(* : coressponding author)

Abstrak−Penyakit batu empedu merupakan salah satu gangguan saluran pencernaan yang paling umum, dan deteksi dini menggunakan data non-invasif dapat mendukung proses penapisan sebelum pemeriksaan pencitraan. Penelitian ini bertujuan meningkatkan kinerja algoritma Decision Tree dalam memprediksi penyakit batu empedu dengan menerapkan metode seleksi fitur Sequential Forward Selection (SFS). Dataset yang digunakan adalah dataset publik Gallstone yang terdiri atas 319 data dan 38 fitur yang mencakup data demografi, bioimpedansi, dan laboratorium. Tahapan penelitian meliputi standardisasi Min-Max, seleksi fitur dengan SFS, serta pembentukan model Decision Tree yang divalidasi menggunakan 10-fold cross validation. Kinerja model dievaluasi menggunakan confusion matrix dengan metrik akurasi dan Area Under the Curve (AUC). Hasil penelitian menunjukkan bahwa Decision Tree tanpa seleksi fitur (38 fitur) hanya mencapai akurasi 67,71% dan AUC 0,677. Setelah penerapan SFS, kinerja terbaik diperoleh dengan hanya menggunakan 6 fitur, yaitu Coronary Artery Disease, Hyperlipidemia, Height, Hepatic Fat Accumulation, C-Reactive Protein, dan Vitamin D, dengan akurasi 76,18% dan AUC 0,762. Dengan demikian, SFS mampu meningkatkan akurasi sebesar 8,46% dan AUC sebesar 0,085 sekaligus mengurangi jumlah fitur sebesar 84%. Temuan ini menunjukkan bahwa SFS tidak hanya meningkatkan kemampuan prediksi Decision Tree, tetapi juga menghasilkan model yang lebih sederhana dan mudah diinterpretasikan untuk prediksi batu empedu.

Kata Kunci: Batu Empedu, *Decision Tree*, Prediksi, Seleksi Fitur, *Sequential Forward Selection*

Abstract−*Gallstone disease is one of the most common gastrointestinal disorders, and early detection using non-invasive data can support screening prior to imaging examinations. This study aims to improve the performance of the Decision Tree algorithm in predicting gallstone disease by applying the Sequential Forward Selection (SFS) feature selection method. The dataset used is the public Gallstone dataset, consisting of 319 records and 38 features covering demographic, bioimpedance, and laboratory data. The research stages comprise Min-Max standardization, feature selection with SFS, and Decision Tree model development validated using 10-fold cross validation. Model performance was evaluated using a confusion matrix with accuracy and Area Under the Curve (AUC) metrics. The results show that the Decision Tree without feature selection (38 features) achieved an accuracy of only 67.71% and an AUC of 0.677. After applying SFS, the best performance was obtained using only 6 features, namely Coronary Artery Disease, Hyperlipidemia, Height, Hepatic Fat Accumulation, C-Reactive Protein, and Vitamin D, with an accuracy of 76.18% and an AUC of 0.762. Thus, SFS increased the accuracy by 8.46% and the AUC by 0.085 while reducing the number of features by 84%. These findings indicate that SFS not only improves the predictive performance of the Decision Tree but also produces a simpler and more interpretable model for gallstone prediction.*

Keywords: *Decision Tree, Feature Selection, Gallstone, Prediction, Sequential Forward Selection*

1. PENDAHULUAN

Penyakit batu empedu (gallstone disease) merupakan salah satu gangguan sistem pencernaan yang paling banyak ditemukan di dunia. Sebuah tinjauan sistematis dan meta-analisis terhadap lebih dari 32 juta partisipan melaporkan prevalensi batu empedu global sebesar 6,1%, dengan angka yang lebih tinggi pada perempuan dibandingkan laki-laki serta meningkat seiring bertambahnya usia (Wang et al., 2024). Penyakit ini juga menimbulkan beban biaya kesehatan yang besar karena dapat berkembang menjadi komplikasi serius seperti kolesistitis, kolangitis, dan pankreatitis apabila tidak terdeteksi lebih awal (Lam et al., 2021). Sejumlah faktor risiko batu empedu telah diidentifikasi, antara lain obesitas, diabetes tipe 2, dan gangguan metabolisme lipid (Yuan et al., 2021; Zhang et al., 2022; Baddam et al., 2023).

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

Diagnosis batu empedu secara konvensional umumnya dilakukan melalui pemeriksaan pencitraan seperti ultrasonografi, CT, atau MRI. Meskipun akurat, teknik pencitraan tersebut memiliki keterbatasan berupa biaya yang relatif tinggi, ketergantungan pada operator, serta keterbatasan ketersediaan alat, sehingga kurang ideal untuk penapisan pada populasi luas (Esen et al., 2024). Oleh karena itu, dibutuhkan pendekatan alternatif yang memanfaatkan data non-pencitraan, seperti data demografi, bioimpedansi, dan laboratorium, untuk membantu mengidentifikasi individu berisiko secara lebih cepat dan ekonomis (Esen et al., 2024).

Perkembangan machine learning membuka peluang besar dalam mendukung diagnosis penyakit berbasis data klinis (Baghdadi et al., 2023; Bhatt et al., 2023). Berbagai algoritma klasifikasi telah banyak diterapkan pada permasalahan medis, dan Decision Tree menjadi salah satu algoritma yang populer karena strukturnya yang sederhana, proses pelatihannya cepat, serta menghasilkan aturan keputusan yang mudah diinterpretasikan oleh tenaga kesehatan (Ozcan & Peker, 2023; Ismafillah et al., 2023). Kemampuan interpretasi ini menjadi nilai penting dalam bidang kesehatan, di mana alasan di balik sebuah prediksi perlu dapat dijelaskan.

Meskipun demikian, kinerja *Decision Tree* sangat dipengaruhi oleh kualitas dan jumlah fitur yang digunakan. Data klinis umumnya memiliki dimensi yang tinggi dan mengandung fitur yang tidak relevan atau redundan. Kehadiran fitur-fitur tersebut dapat menurunkan akurasi, memperbesar kompleksitas model, dan meningkatkan risiko *overfitting.* Untuk mengatasi permasalahan ini, seleksi fitur (*feature selection*) menjadi tahap prapemrosesan penting yang bertujuan memilih subset fitur paling informatif sehingga dapat meningkatkan efisiensi dan akurasi model pembelajaran (Aswal et al., 2023; Alija et al., 2023).

Sequential Forward Selection (SFS) merupakan salah satu metode seleksi fitur berbasis wrapper yang bekerja secara bottom-up (Chotchantarakun, 2023). Metode ini dimulai dari himpunan kosong, kemudian menambahkan satu fitur pada setiap iterasi yang paling meningkatkan kinerja model, hingga diperoleh subset fitur terbaik. SFS dikenal mudah dipahami, efisien secara komputasi, dan secara langsung mengoptimalkan metrik kinerja model yang digunakan sebagai kriteria seleksi (Aswal et al., 2023). Karakteristik ini menjadikan SFS cocok dipadukan dengan Decision Tree untuk menemukan kombinasi fitur yang paling relevan.

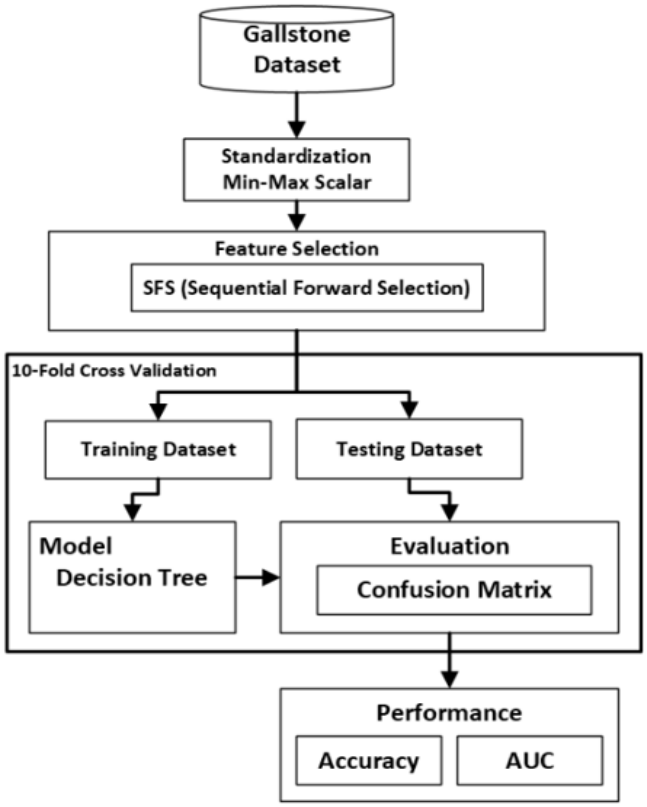
Dataset Gallstone yang bersifat publik menyediakan 319 data dengan 38 fitur (Esen et al., 2024), sehingga jumlah fitur yang cukup banyak berpotensi menurunkan kinerja model apabila digunakan seluruhnya. Beberapa penelitian sebelumnya pada dataset ini berfokus pada perbandingan berbagai algoritma untuk mencapai akurasi setinggi mungkin (Esen et al., 2024; Chakraborty & Mukherjee, 2025), namun aspek penyederhanaan model melalui seleksi fitur pada Decision Tree masih belum banyak dieksplorasi secara khusus. Hal inilah yang menjadi celah penelitian yang diangkat pada studi ini.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk meningkatkan kinerja algoritma *Decision Tree* dalam memprediksi penyakit batu empedu dengan menerapkan seleksi fitur *Sequential Forward Selection*. Kontribusi penelitian ini adalah menunjukkan sejauh mana SFS mampu meningkatkan akurasi dan AUC *Decision Tree* sekaligus mengurangi jumlah fitur, sehingga dihasilkan model prediksi yang lebih ringkas, akurat, dan mudah diinterpretasikan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen dengan membandingkan kinerja *Decision Tree* tanpa seleksi fitur dan dengan seleksi fitur *Sequential Forward Selection*. Secara garis besar, tahapan penelitian terdiri atas pengumpulan dataset, standardisasi data dengan Min-Max, seleksi fitur menggunakan SFS, pembentukan model *Decision Tree* yang divalidasi dengan 10-fold cross validation, serta evaluasi kinerja menggunakan *confusion matrix* dengan metrik akurasi dan AUC. Alur penelitian secara lengkap ditunjukkan pada Gambar 1.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815



Gambar 1. Rancangan Model FS-DT untuk Prediksi Batu Empedu

2.1 Dataset

Dataset yang digunakan adalah dataset publik Gallstone yang tersedia pada UCI *Machine Learning Repository*. Dataset ini terdiri atas 319 data individu, di mana 161 di antaranya terdiagnosis menderita batu empedu dan sisanya merupakan kelompok kontrol sehat, sehingga distribusi kelasnya relatif seimbang. Dataset memiliki 38 fitur yang mencakup data demografi (antara lain usia, jenis kelamin, tinggi badan, berat badan, dan indeks massa tubuh), data bioimpedansi (antara lain total body water, lean mass, hepatic fat accumulation, dan visceral fat area), serta data laboratorium (antara lain glukosa, kolesterol total, LDL, HDL, trigliserida, C-Reactive Protein, dan vitamin D). Variabel target berupa status batu empedu yang bersifat biner, yaitu menderita atau tidak menderita batu empedu (Esen et al., 2024). Dataset ini tidak memiliki nilai yang hilang sehingga tidak memerlukan penanganan missing value tambahan. Dataset yang sama juga telah dimanfaatkan pada penelitian prediksi batu empedu berbasis *machine learning* lainnya (Chakraborty & Mukherjee, 2025).

2.2 Standardisasi *Min-Max*

Fitur pada dataset memiliki rentang nilai yang sangat beragam, misalnya nilai laboratorium dan nilai bioimpedansi berada pada skala yang berbeda jauh. Perbedaan skala ini dapat menyebabkan fitur bernilai besar mendominasi proses pembelajaran. Oleh karena itu, dilakukan standardisasi menggunakan metode Min-Max Scaler yang mentransformasikan seluruh nilai fitur ke dalam rentang 0 sampai 1. Transformasi Min-Max dihitung dengan membagi selisih antara nilai fitur dan nilai minimum terhadap selisih antara nilai maksimum dan minimum pada fitur tersebut. Dengan cara ini, setiap fitur memiliki kontribusi yang lebih seimbang dan proses pelatihan model menjadi lebih stabil. Standardisasi Min-Max telah terbukti berpengaruh terhadap peningkatan akurasi berbagai algoritma machine learning (Cabello-Solorzano et al., 2023; Shantal et al., 2023).

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

2.3 *Sequential Forward Selection*

Sequential Forward Selection (SFS) merupakan metode seleksi fitur berbasis wrapper yang menelusuri subset fitur secara bertahap (Chotchantarakun, 2023). Proses SFS dimulai dari himpunan fitur kosong. Pada setiap iterasi, algoritma mengevaluasi penambahan setiap fitur yang belum terpilih dan memilih satu fitur yang memberikan peningkatan kinerja model paling besar. Fitur terpilih kemudian dimasukkan ke dalam subset dan tidak dievaluasi ulang pada iterasi berikutnya. Proses ini diulang hingga seluruh fitur diuji, sehingga diperoleh urutan penambahan fitur beserta kinerja model pada setiap jumlah fitur. Pada penelitian ini, kinerja Decision Tree digunakan sebagai kriteria seleksi, dan jumlah fitur yang menghasilkan kinerja terbaik dipilih sebagai subset fitur akhir. Pendekatan seleksi fitur berbasis *wrapper* juga telah terbukti meningkatkan kinerja klasifikasi pada berbagai kasus medis (Kurnia et al., 2023; Hidayat et al., 2023).

2.4 *Decision Tree*

Decision Tree adalah algoritma klasifikasi yang membentuk model berupa struktur pohon, terdiri atas simpul akar, simpul internal, dan simpul daun. Setiap simpul internal merepresentasikan pengujian terhadap sebuah fitur, setiap cabang merepresentasikan hasil pengujian, dan setiap simpul daun merepresentasikan label kelas. Pembentukan pohon dilakukan dengan memilih fitur pemisah terbaik pada setiap simpul berdasarkan ukuran kemurnian seperti indeks Gini atau information gain (Ozcan & Peker, 2023). *Decision Tree* dipilih pada penelitian ini karena mampu menghasilkan aturan keputusan yang transparan dan mudah diinterpretasikan, sehingga relevan untuk mendukung pengambilan keputusan pada domain kesehatan (Ismafillah et al., 2023).

2.5 Validasi dan Evaluasi

Model dievaluasi menggunakan 10-fold cross validation, di mana dataset dibagi menjadi sepuluh bagian secara bergantian sebagai data latih dan data uji. Skema ini memberikan estimasi kinerja yang lebih adil dan mengurangi risiko bias akibat pembagian data tunggal. Hasil klasifikasi dirangkum dalam *confusion matrix* yang memuat nilai *True Positive, True Negative, False Positive*, dan *False Negative.* Dari *confusion matrix* tersebut dihitung metrik akurasi, yaitu proporsi prediksi yang benar terhadap seluruh data, serta nilai *Area Under the Curve* (AUC) yang menggambarkan kemampuan model dalam membedakan kelas positif dan negatif. Kedua metrik ini digunakan sebagai dasar perbandingan kinerja *Decision Tree* sebelum dan sesudah penerapan SFS.

3. ANALISA DAN PEMBAHASAN

3.1 Kinerja *Decision Tree* tanpa Seleksi Fitur

Pada tahap awal, Decision Tree dibangun menggunakan seluruh 38 fitur tanpa penerapan seleksi fitur. Hasil pengujian menunjukkan bahwa model hanya mencapai akurasi sebesar 67,71% dengan nilai AUC 0,677. Nilai ini mengindikasikan bahwa penggunaan seluruh fitur tidak serta-merta menghasilkan kinerja yang optimal, karena sebagian fitur kemungkinan bersifat tidak relevan atau redundan sehingga justru menambah kompleksitas model tanpa memberikan kontribusi informasi yang berarti. Kondisi ini menjadi dasar penerapan seleksi fitur untuk menyaring fitur yang benar-benar informatif.

3.2 Hasil Penerapan *Sequential Forward Selection*

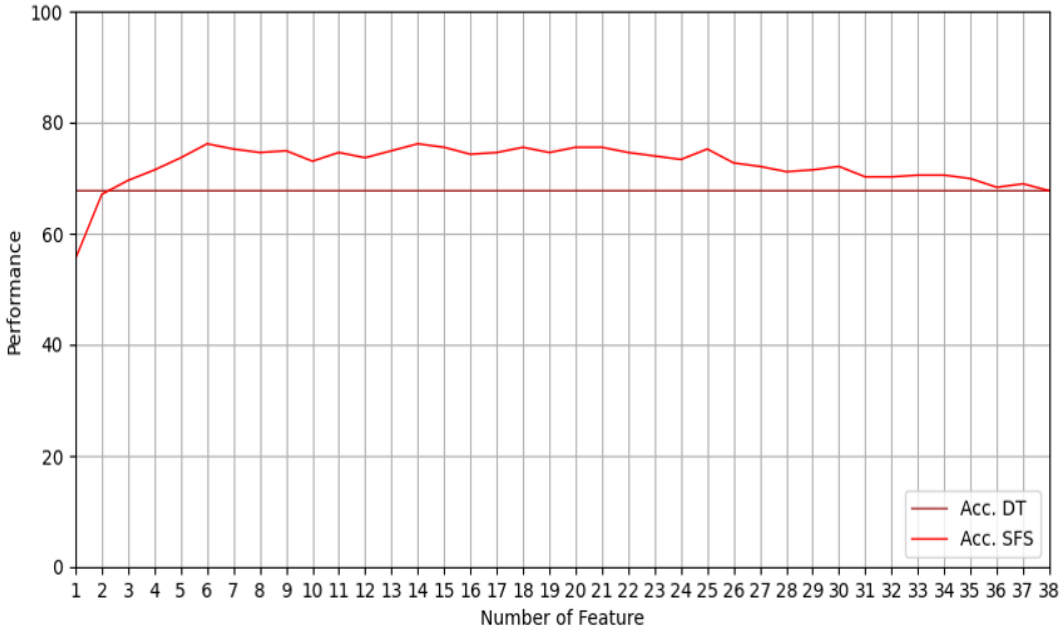
Penerapan SFS dilakukan dengan mengevaluasi kinerja Decision Tree pada setiap penambahan jumlah fitur, mulai dari 1 fitur hingga 38 fitur. Rangkuman kinerja untuk setiap jumlah fitur ditunjukkan pada Tabel 1. Terlihat bahwa kinerja model meningkat tajam pada penambahan fitur awal, kemudian mencapai puncak, dan cenderung menurun kembali ketika terlalu banyak fitur ditambahkan. Kinerja terbaik dicapai pada penggunaan 6 fitur dengan akurasi 76,18% dan AUC 0,762.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

Tabel 1. Kinerja *Decision Tree* pada Setiap Jumlah Fitur Hasil SFS

Fitur	Akurasi (%)	AUC	Fitur	Akurasi (%)	AUC
1	55,49	0,553	20	75,55	0,756
2	67,08	0,671	21	75,55	0,755
3	69,59	0,696	22	74,61	0,746
4	71,47	0,715	23	73,98	0,740
5	73,67	0,737	24	73,35	0,734
6	76,18	0,762	25	75,24	0,752
7	75,24	0,752	26	72,73	0,727
8	74,61	0,746	27	72,10	0,721
9	74,92	0,749	28	71,16	0,712
10	73,04	0,730	29	71,47	0,715
11	74,61	0,746	30	72,10	0,721
12	73,67	0,737	31	70,22	0,702
13	74,92	0,749	32	70,22	0,702
14	76,18	0,762	33	70,53	0,705
15	75,55	0,755	34	70,53	0,705
16	74,29	0,743	35	69,91	0,699
17	74,61	0,746	36	68,34	0,683
18	75,55	0,756	37	68,97	0,690
19	74,61	0,747	38	67,71	0,677

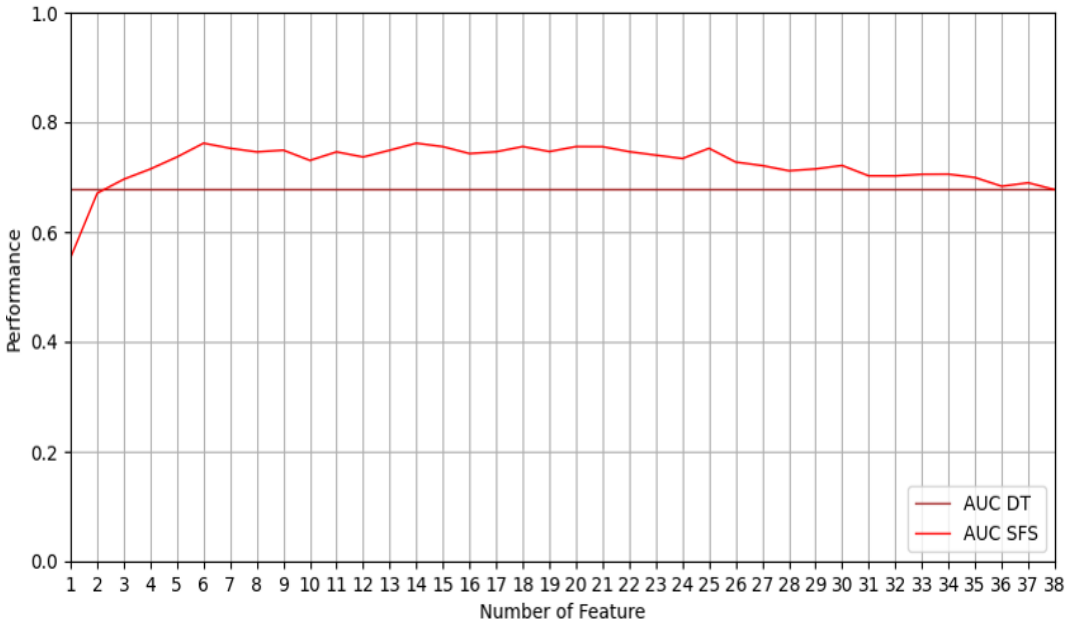
Kecenderungan tersebut juga terlihat jelas pada grafik kinerja. Gambar 2 menampilkan perbandingan akurasi *Decision Tree* pada setiap jumlah fitur (Acc. DT) terhadap garis akurasi *Decision Tree* yang menggunakan seluruh fitur sebagai acuan (Acc. SFS). Akurasi meningkat cepat pada fitur pertama hingga keenam, mencapai puncak, lalu berangsur menurun mendekati garis acuan ketika jumlah fitur bertambah banyak.



Gambar 2. Grafik Akurasi *Decision Tree* pada Setiap Jumlah Fitur

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

Pola serupa ditemukan pada nilai AUC seperti ditunjukkan pada Gambar 3. Nilai AUC tertinggi juga dicapai pada penggunaan 6 fitur, sejalan dengan hasil akurasi. Konsistensi antara akurasi dan AUC memperkuat kesimpulan bahwa subset 6 fitur merupakan konfigurasi paling optimal bagi model.



Gambar 3. Grafik AUC *Decision Tree* pada Setiap Jumlah Fitur

3.3 Model Terbaik dan Fitur Terpilih

Berdasarkan hasil SFS, model terbaik diperoleh dengan hanya menggunakan 6 fitur, yaitu *Coronary Artery Disease* (CAD), *Hyperlipidemia, Height, Hepatic Fat Accumulation* (HFA), C-*Reactive Protein* (CRP), dan Vitamin D. Dengan subset fitur ini, model mencapai akurasi 76,18% dan AUC 0,762. *Confusion matrix* dari model terbaik ditunjukkan pada Tabel 2, dengan total 319 data uji.

Tabel 2. *Confusion Matrix* Model Terbaik (6 fitur)

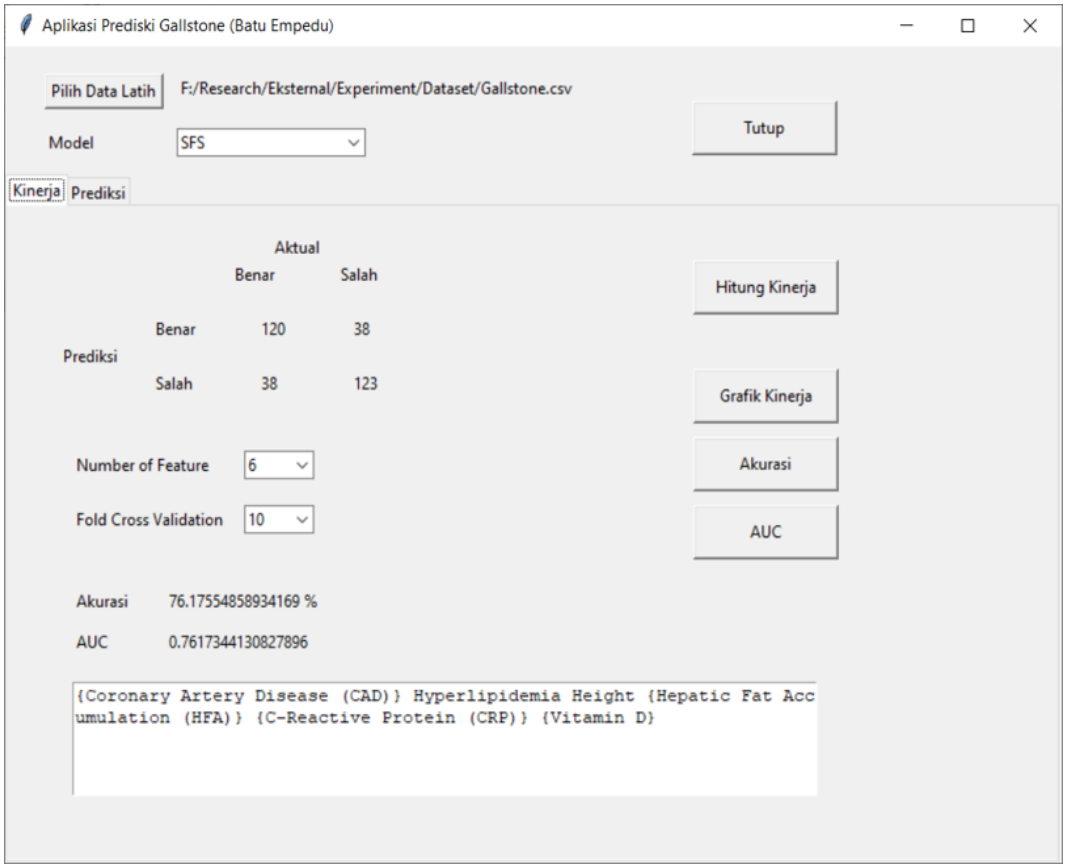
	Aktual Benar	**Aktual Salah**
Prediksi Benar	120	38
Prediksi Salah	38	123

Dari confusion matrix tersebut, model menghasilkan 120 *True Positive*, 123 *True Negative*, 38 *False Positive*, dan 38 *False Negative*. Nilai ini setara dengan akurasi sebesar 76,18%, dengan presisi dan recall yang seimbang pada kisaran 75,95%, serta spesifisitas sekitar 76,40%. Keseimbangan antara presisi dan recall menunjukkan bahwa model tidak berat sebelah dalam mengenali pasien yang menderita maupun yang tidak menderita batu empedu.

3.4 Implementasi Aplikasi

Untuk memudahkan pengujian, model diimplementasikan ke dalam sebuah aplikasi prediksi batu empedu. Aplikasi ini memungkinkan pengguna memilih data latih, menentukan model SFS, mengatur jumlah fitur serta jumlah fold cross validation, kemudian menghitung kinerja berupa confusion matrix, akurasi, dan AUC. Tampilan aplikasi beserta hasil kinerja pada konfigurasi 6 fitur ditunjukkan pada Gambar 4.

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815



Gambar 4. Tampilan Aplikasi Prediksi Batu Empedu dengan Model SFS 6 Fitur

3.5 Pembahasan

Hasil penelitian menunjukkan bahwa penerapan *Sequential Forward Selection* memberikan peningkatan kinerja yang nyata pada *Decision Tree*. Akurasi meningkat dari 67,71% menjadi 76,18%, atau naik sebesar 8,46%, sedangkan AUC meningkat dari 0,677 menjadi 0,762, atau naik sebesar 0,085. Peningkatan ini dicapai sekaligus dengan pengurangan jumlah fitur dari 38 menjadi hanya 6 fitur, yang berarti terjadi reduksi dimensi sebesar sekitar 84%. Rangkuman perbandingan kinerja disajikan pada Tabel 3.

Tabel 3. Perbandingan Kinerja *Decision Tree* Sebelum dan Sesudah SFS

Model	Jumlah Fitur	Akurasi (%)	AUC
Decision Tree (tanpa seleksi fitur)	38	67,71	0,677
Decision Tree + SFS (usulan)	6	76,18	0,762

Temuan ini menegaskan bahwa penggunaan seluruh fitur tidak selalu menghasilkan model terbaik. Fitur yang tidak relevan dan redundan dapat menambah derau serta memperbesar kompleksitas pohon keputusan sehingga menurunkan kemampuan generalisasi. Dengan menyaring fitur melalui SFS, model menjadi lebih fokus pada informasi yang benar-benar diskriminatif, sehingga kinerjanya meningkat sekaligus lebih ringkas. Peningkatan kinerja akibat seleksi fitur ini sejalan dengan temuan pada penelitian klasifikasi medis lainnya (Kurnia et al., 2023; Hidayat et al., 2023).

Fitur-fitur yang terpilih juga selaras dengan temuan pada literatur medis dan penelitian *machine learning* sebelumnya pada dataset yang sama. C-*Reactive Protein* dan Vitamin D,

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

misalnya, secara konsisten dilaporkan sebagai penanda penting yang berkaitan dengan pembentukan batu empedu (Esen et al., 2024; Chakraborty & Mukherjee, 2025), sementara Hyperlipidemia dan Coronary Artery Disease berkaitan erat dengan gangguan metabolisme lipid yang menjadi salah satu faktor risiko (Zhang et al., 2022; Yuan et al., 2021). Kesesuaian ini memperkuat validitas hasil seleksi fitur, karena subset fitur yang dihasilkan tidak hanya meningkatkan kinerja secara statistik, tetapi juga masuk akal secara klinis.

Dibandingkan dengan penelitian terdahulu yang menekankan pencapaian akurasi tertinggi melalui algoritma ensemble yang kompleks (Esen et al., 2024), penelitian ini menawarkan sudut pandang berbeda, yaitu menghasilkan model yang sederhana, cepat, dan mudah diinterpretasikan tanpa mengorbankan peningkatan kinerja. Upaya optimasi kinerja model klasifikasi berbasis pohon keputusan juga menjadi perhatian pada berbagai kasus prediksi penyakit lainnya (Bhatt et al., 2023; Chandrasekhar & Peddakrishna, 2023; Nicholas, 2025). Model dengan 6 fitur jauh lebih praktis untuk diterapkan pada proses penapisan awal, karena hanya memerlukan sedikit variabel pemeriksaan. Meskipun demikian, penelitian ini memiliki keterbatasan, antara lain penggunaan satu dataset dengan jumlah data yang relatif terbatas serta fokus pada satu algoritma klasifikasi.

4. KESIMPULAN

4.1 Kesimpulan

Penelitian ini menerapkan seleksi fitur *Sequential Forward Selection* untuk meningkatkan kinerja Decision Tree dalam prediksi penyakit batu empedu menggunakan dataset publik Gallstone. Hasil penelitian menunjukkan bahwa Decision Tree tanpa seleksi fitur dengan 38 fitur hanya mencapai akurasi 67,71% dan AUC 0,677. Setelah penerapan SFS, kinerja terbaik diperoleh dengan hanya 6 fitur, yaitu Coronary Artery Disease, Hyperlipidemia, Height, Hepatic Fat Accumulation, C-Reactive Protein, dan Vitamin D, dengan akurasi 76,18% dan AUC 0,762. Dengan demikian, SFS mampu meningkatkan akurasi sebesar 8,46% dan AUC sebesar 0,085 sekaligus mengurangi jumlah fitur hingga sekitar 84%. Hal ini membuktikan bahwa SFS efektif menghasilkan model prediksi yang lebih akurat, ringkas, dan mudah diinterpretasikan sehingga berpotensi mendukung penapisan dini batu empedu berbasis data non-pencitraan.

4.2 Saran

Penelitian selanjutnya dapat mengembangkan studi ini dengan membandingkan SFS terhadap metode seleksi fitur lain seperti *Sequential Backward Selection*, algoritma berbasis metaheuristik, maupun metode filter, serta menerapkannya pada algoritma klasifikasi lain untuk menguji konsistensi peningkatan kinerja. Selain itu, pengujian pada dataset yang lebih besar dan lebih beragam serta penambahan optimasi *hyperparameter Decision Tree* dapat dilakukan untuk memperoleh model prediksi batu empedu yang lebih andal dan dapat digeneralisasi.

REFERENCES

Alija, S., Beqiri, E., Gaafar, A. S., & Hamoud, A. K. (2023). Predicting Students Performance Using Supervised Machine Learning Based on Imbalanced Dataset and Wrapper Feature Selection. Informatica, 47(1), 11-20.

Aswal, S., Jyothi, A., & Mehra, R. (2023). Feature Selection Method Based on Honeybee-SMOTE for Medical Data Classification. Informatica, 46(9), 111-118. https://doi.org/10.31449/inf.v46i9.4098

Baddam, A., Akuma, O., Raj, R., Akuma, C. M., Augustine, S. W., & Sheikh Hanafi, I. (2023). Analysis of Risk Factors for Cholelithiasis: A Single-Center Retrospective Study. Cureus, 15(9), e46155. https://doi.org/10.7759/cureus.46155

OKTAL : Jurnal Ilmu Komputer dan Science
Volume 5, No. 8, Agustus 2026
ISSN 2828-2442 (media online)
Hal 807-815

Baghdadi, N. A., Farghaly Abdelaliem, S. M., Malki, A., Gad, I., Ewis, A., & Atlam, E. (2023). Advanced machine learning techniques for cardiovascular disease early detection and diagnosis. Journal of Big Data, 10(1), 144. https://doi.org/10.1186/s40537-023-00817-1

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms, 16(2), 88. https://doi.org/10.3390/a16020088

Cabello-Solorzano, K., Ortigosa de Araujo, I., Pena, M., Correia, L., & Tallon-Ballesteros, A. J. (2023). The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis. In Proceedings of the 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023) (Vol. 750, pp. 344-353). Springer. https://doi.org/10.1007/978-3-031-42536-3_33

Chakraborty, C., & Mukherjee, N. (2025). Bayesian Hybrid Machine Learning of Gallstone Risk. arXiv preprint arXiv:2506.14561. https://doi.org/10.48550/arXiv.2506.14561

Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization. Processes, 11(4), 1210. https://doi.org/10.3390/pr11041210

Chotchantarakun, K. (2023). Optimizing Sequential Forward Selection on Classification using Genetic Algorithm. Informatica, 47(9), 1-12. https://doi.org/10.31449/inf.v46i9.4964

Esen, I., Arslan, H., Akturk Esen, S., Gulsen, M., Kultekin, N., & Ozdemir, O. (2024). Early prediction of gallstone disease with a machine learning-based method from bioimpedance and laboratory data. Medicine, 103(8), e37258. https://doi.org/10.1097/MD.0000000000037258

Hidayat, R., Kartini, D., Mazdadi, M. I., Budiman, I., & Ramadhani, R. (2023). Implementasi Seleksi Fitur Binary Particle Swarm Optimization pada Algoritma K-NN untuk Klasifikasi Kanker Payudara. Jurnal Sistem dan Teknologi Informasi, 11(2), 135-139.

Ismafillah, D., Rohana, T., & Cahyana, Y. (2023). Analisis Algoritma Pohon Keputusan untuk Memprediksi Penyakit Diabetes Menggunakan Oversampling SMOTE. INFOTECH: Jurnal Informatika dan Teknologi, 4(1), 27-36.

Kurnia, D., Mazdadi, M. I., Kartini, D., Nugroho, R. A., & Abadi, F. (2023). Seleksi Fitur dengan Particle Swarm Optimization pada Klasifikasi Penyakit Parkinson Menggunakan XGBoost. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 10(5), 1083-1094.

Lam, R., Zakko, A., Petrov, J. C., Kumar, P., Duffy, A. J., & Muniraj, T. (2021). Gallbladder Disorders: A Comprehensive Review. Disease-a-Month, 67(7), 101130. https://doi.org/10.1016/j.disamonth.2021.101130

Nicholas. (2025). Heart Disease Prediction with Decision Tree. Social Science and Humanities Journal, 9(1), 6451-6457. https://doi.org/10.18535/sshj.v9i01.1444

Ozcan, M., & Peker, S. (2023). A classification and regression tree algorithm for heart disease modeling and prediction. Healthcare Analytics, 3, 100130. https://doi.org/10.1016/j.health.2022.100130

Shantal, M., Othman, Z., & Abu Bakar, A. (2023). A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min-Max Normalization. Symmetry, 15(12), 2185. https://doi.org/10.3390/sym15122185

Wang, X., Yu, W., Jiang, G., Li, H., Li, S., Xie, L., Bai, X., Cui, P., Chen, Q., Lou, Y., Zou, L., Li, S., Zhou, Z., Zhang, C., Sun, P., & Mao, M. (2024). Global Epidemiology of Gallstones in the 21st Century: A Systematic Review and Meta-Analysis. Clinical Gastroenterology and Hepatology, 22(8), 1586-1595. https://doi.org/10.1016/j.cgh.2024.01.051

Yuan, S., Gill, D., Giovannucci, E. L., & Larsson, S. C. (2021). Obesity, Type 2 Diabetes, Lifestyle Factors, and Risk of Gallstone Disease: A Mendelian Randomization Investigation. Clinical Gastroenterology and Hepatology, 19(12), 2540-2548.

Zhang, M., Mao, M., Zhang, C., Hu, F., Cui, P., Li, G., Shi, J., Wang, X., & Shan, X. (2022). Blood lipid metabolism and the risk of gallstone disease: A multi-center cross-sectional study and meta-analysis. Lipids in Health and Disease, 21(1), 26. https://doi.org/10.1186/s12944-022-01635-9