

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Gagal jantung (*Heart Failure/HF*) merupakan salah satu penyakit kardiovaskular yang menjadi penyebab utama tingginya angka kesakitan dan kematian di berbagai negara. Meskipun telah terjadi perkembangan dalam diagnosis dan terapi, pasien gagal jantung masih memiliki tingkat rawat inap ulang (*readmission*) yang tinggi, khususnya dalam periode 30 hari setelah pasien dipulangkan, dengan kisaran sekitar 20%–27% di Amerika Utara [1]. Kejadian *readmission* tidak hanya berdampak pada kondisi kesehatan dan kualitas hidup pasien, tetapi juga meningkatkan beban biaya layanan kesehatan secara signifikan. Di Canada, sekitar satu dari sebelas pasien gagal jantung mengalami rawat inap ulang yang menyebabkan beban biaya kesehatan melebihi 2,3 miliar dolar Kanada per tahun [2].

Pemilihan horizon prediksi selama 30 hari didasarkan pada dua pertimbangan. Pertama, secara klinis *30-day readmission* telah digunakan sebagai salah satu indikator untuk mengevaluasi kualitas pelayanan rumah sakit pada pasien gagal jantung, karena kejadian rawat inap ulang dalam periode tersebut mencerminkan keberhasilan transisi perawatan dan manajemen pasien setelah dipulangkan [3]. Kedua, sebagian besar penelitian terdahulu juga menggunakan horizon prediksi selama 30 hari sehingga hasil penelitian ini dapat dibandingkan secara langsung. Meskipun demikian, performa model pada penelitian-penelitian tersebut masih menunjukkan hasil yang beragam, sehingga prediksi *30-day readmission* masih menjadi tantangan dan memerlukan pengembangan model dengan performa yang lebih baik.

Dalam praktik klinis, prediksi risiko *readmission* masih menjadi permasalahan yang kompleks. Hal ini disebabkan oleh banyaknya faktor yang saling berinteraksi, seperti kondisi klinis pasien, biomarker laboratorium, riwayat perawatan, serta faktor demografis [4]. Pendekatan konvensional yang digunakan oleh tenaga medis umumnya masih bergantung pada pengalaman klinis dan aturan sederhana sehingga memiliki keterbatasan dalam mengidentifikasi pola yang kompleks serta hubungan non-linear di antara variabel-variabel pasien [5].

Untuk mengatasi permasalahan tersebut, pendekatan *machine learning*

banyak dikembangkan karena mampu mempelajari hubungan yang kompleks dan non-linear antarvariabel klinis yang sulit ditangkap oleh metode statistik konvensional. Dengan memanfaatkan data rekam medis elektronik, model *machine learning* dapat membantu mengidentifikasi pasien yang berisiko tinggi mengalami *readmission* sehingga intervensi pencegahan dapat dilakukan lebih dini sebelum pasien dipulangkan. Berbagai penelitian sebelumnya telah menerapkan algoritma seperti *Logistic Regression*, *Random Forest*, dan *XGBoost* untuk memprediksi *readmission* pasien gagal jantung. Namun, performa yang diperoleh masih menunjukkan hasil yang beragam sehingga kemampuan prediksi *30-day readmission* belum sepenuhnya optimal. Selain itu, sebagian besar penelitian masih berfokus pada penggunaan model tunggal sehingga kemampuan generalisasi model masih berpotensi untuk ditingkatkan. Salah satu pendekatan yang dapat mengatasi keterbatasan tersebut adalah *stacked ensemble*, yang menggabungkan keunggulan beberapa algoritma dasar melalui sebuah *meta learner* untuk menghasilkan prediksi yang lebih akurat dan stabil.

Selain performa prediksi, interpretabilitas model juga menjadi aspek penting dalam implementasi di bidang kesehatan. Hasil prediksi perlu dapat dijelaskan secara klinis agar dapat dipercaya dan digunakan sebagai dasar pengambilan keputusan oleh tenaga medis. Di sisi lain, implementasi model prediksi di lingkungan rumah sakit tidak hanya ditentukan oleh akurasi, tetapi juga oleh efisiensi biaya pemeriksaan klinis. Dalam praktiknya, berbagai parameter laboratorium seperti *Blood Urea Nitrogen* (BUN), *creatinine*, *sodium*, dan *hemoglobin* relatif memiliki biaya pemeriksaan yang rendah, sedangkan parameter seperti *NT-proBNP* serta pengukuran *ejection fraction* melalui *echocardiography* memerlukan biaya yang lebih tinggi [6]. Meskipun parameter-parameter tersebut penting secara klinis, tidak semua memberikan kontribusi yang signifikan terhadap peningkatan performa model prediksi.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada pengembangan model dengan metode *ensemble learning* untuk prediksi *30-day readmission* pasien gagal jantung menggunakan pendekatan *stacked ensemble* yang mengombinasikan algoritma *Random Forest*, *XGBoost*, dan *Logistic Regression*. Model *stacked ensemble* kemudian dibandingkan dengan performa masing-masing algoritma individual untuk mengevaluasi efektivitas pendekatan yang diusulkan. Selain itu, penelitian ini juga menerapkan analisis interpretabilitas menggunakan SHAP serta melakukan seleksi fitur untuk mengidentifikasi kombinasi variabel klinis yang mampu mempertahankan performa model dengan jumlah fitur yang

lebih sedikit. Melalui pendekatan tersebut, penelitian ini diharapkan dapat menghasilkan model prediksi yang lebih akurat, mudah diinterpretasikan, serta lebih efisien untuk diterapkan dalam praktik klinis.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang tersebut, permasalahan dalam penelitian ini tidak hanya berfokus pada peningkatan akurasi prediksi, tetapi juga pada efisiensi biaya dalam pemanfaatan fitur-fitur klinis. Oleh karena itu, rumusan masalah penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana membangun model prediksi *30-day readmission* menggunakan pendekatan *ensemble learning*?
2. Algoritma manakah yang memberikan performa prediksi terbaik terhadap *30-day readmission* pasien gagal jantung?
3. Apakah metode *stacked ensemble* dapat meningkatkan performa dibandingkan model tunggal?
4. Faktor-faktor klinis apa saja yang paling berpengaruh terhadap risiko *readmission* pasien?
5. Bagaimana interpretasi hasil model menggunakan *explainable AI* dapat mendukung pengambilan keputusan klinis?

1.3 Batasan Permasalahan

Agar penelitian ini lebih terfokus serta memiliki cakupan yang jelas, maka ditentukan beberapa batasan penelitian sebagai berikut:

1. Dataset yang digunakan merupakan data sekunder klinis pasien penyakit jantung dengan tipe data tabular.
2. Prediksi yang dilakukan terbatas pada *readmission* dalam waktu 30 hari setelah pasien dipulangkan.
3. Variabel yang digunakan meliputi data demografi, riwayat perawatan, vital signs, serta biomarker laboratorium.

4. Algoritma yang digunakan meliputi *Logistic Regression* sebagai baseline, *Random Forest*, *XGBoost*, *LightGBM*, dan *Stacked Ensemble*.
5. Evaluasi model menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-score*, dan *ROC-AUC*.
6. Penelitian tidak membahas implementasi sistem secara langsung di rumah sakit, melainkan berfokus pada pengembangan dan evaluasi model.
7. Validasi model dilakukan menggunakan metode cross-validation

1.4 Tujuan Penelitian

Mengacu pada rumusan masalah yang telah disusun, penelitian ini bertujuan untuk membangun model prediksi yang memiliki tingkat akurasi tinggi sekaligus mempertimbangkan efisiensi biaya. Adapun tujuan penelitian ini secara lebih spesifik adalah sebagai berikut:

1. Mengembangkan model *machine learning* untuk memprediksi *readmission* pasien gagal jantung.
2. Membandingkan performa berbagai algoritma *machine learning* dalam prediksi *readmission*.
3. Mengimplementasikan metode *stacked ensemble* untuk meningkatkan performa model.
4. Mengidentifikasi faktor-faktor penting yang mempengaruhi risiko *readmission* pasien.
5. Menginterpretasikan hasil model menggunakan pendekatan *explainable AI*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi baik dari sisi akademik maupun praktis, serta menghasilkan manfaat ekonomi yang mendukung penerapannya dalam layanan kesehatan.

1.5.1 Manfaat Akademis

- Memberikan kontribusi dalam pengembangan metode *machine learning* berbasis data klinis.
- Menjadi referensi bagi penelitian selanjutnya terkait prediksi readmission dan *cost-aware machine learning*.

1.5.2 Manfaat Praktis

- Membantu tenaga medis dalam mengidentifikasi pasien dengan risiko readmission secara lebih akurat.
- Mendukung pengambilan keputusan klinis berbasis data dan interpretasi model.
- Memberikan alternatif strategi pemeriksaan klinis yang lebih efisien.

1.5.3 Manfaat Ekonomis

- Mengurangi biaya pemeriksaan laboratorium dengan mengidentifikasi fitur yang paling relevan.
- Membantu rumah sakit dalam mengoptimalkan penggunaan sumber daya medis.
- Mendukung penerapan layanan kesehatan yang lebih efisien dan berkelanjutan.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Menguraikan latar belakang masalah, rumusan masalah, batasan penelitian, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

- Bab 2 LANDASAN TEORI

Membahas teori-teori yang mendukung penelitian, termasuk machine learning, algoritma klasifikasi, ensemble learning, serta explainable AI.

- Bab 3 METODOLOGI PENELITIAN

Menjelaskan tahapan penelitian yang meliputi pengumpulan data, preprocessing, feature engineering, pembagian data, pemodelan, serta evaluasi model.

- Bab 4 HASIL DAN DISKUSI

Menyajikan hasil eksperimen, analisis performa model, perbandingan antar model, serta interpretasi hasil.

- Bab 5 KESIMPULAN DAN SARAN

Berisi kesimpulan dari penelitian serta saran untuk penelitian selanjutnya.

