

BAB I

PENDAHULUAN

1.1 Latar Belakang

Gagal jantung adalah salah satu penyebab utama kematian, baik di Indonesia maupun di berbagai negara lainnya, dan sering kali menjadi faktor utama pasien di ICU (*Intensive Care Unit*) mengalami kondisi kritis [1]. Pasien yang berada di ICU yang mengalami penyakit gagal jantung sering mengalami komplikasi yang cukup serius yang nantinya dapat menyebabkan kematian jika pasien tidak segera ditangani oleh pihak medis dengan cepat dan tepat [2]. Berdasarkan data dari *Mayo Clinic* terhadap 4.012 pasien yang dirawat di *Cardiac Intensive Care Unit* (CICU) selama periode 2007–2018, tercatat angka kematian di rumah sakit mencapai 12,5% [3]. Studi tahun 2023 menganalisis data dari 7.202 pasien lansia dengan gagal jantung yang dirawat di ICU di Amerika Serikat. Hasil penelitian menunjukkan bahwa tingkat kematian dalam 30 hari mencapai 16%, dengan faktor risiko utama seperti denyut jantung tinggi, tekanan darah sistolik rendah, kadar NT-*proBNP* tinggi, dan gangguan fungsi ginjal [4]. Lalu data tahun 2019, angka kematian di dunia yang disebabkan oleh gagal jantung mencapai 18,6 juta [5]. Fenomena kematian yang disebabkan oleh gagal jantung ini juga terjadi di Indonesia, Kementerian Kesehatan Republik Indonesia melaporkan bahwa penyakit jantung bertanggung jawab atas lebih dari 31% dari seluruh kematian di negara ini pada tahun 2020, menjadikannya penyebab utama kematian [6]. Lalu, pada tahun 2019 laporan dari Kementerian Kesehatan RI mencatat bahwa pada tahun 2019 para petugas pemilu banyak yang meninggal dunia, yang dikarenakan oleh penyakit jantung. Total dari jumlah meninggal dunia tersebut ada sebanyak 48 yang dilaporkan [7]. Dari fenomena tersebut, terlihat bahwa penyakit gagal jantung merupakan kondisi yang sangat serius. Oleh karena itu, melakukan deteksi dini dan prediksi kemungkinan terjadinya gagal jantung pada pasien ICU merupakan langkah yang dapat dilakukan untuk mencegah kematian.

Pada era ini, perkembangan AI (*Artificial Intelligence*) dan ML (*Machine Learning*) telah memberikan perubahan besar dalam membantu dunia medis dalam

menganalisis data [8]. Data medis dalam jumlah besar dan kompleks dapat dianalisis menggunakan ML untuk mengidentifikasi pola-pola yang sulit dipahami secara manual [8]. Terdapat berbagai model ML yang dapat digunakan untuk melakukan prediksi gagal jantung berdasarkan data ICU, tetapi efektivitas model dapat berbeda tergantung pada kompleksitas data, parameter yang digunakan, dan metode evaluasi yang diterapkan. Model seperti *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Logistic Regression* (LR) sering digunakan untuk mengklasifikasikan penyakit [9][10][11].

Random Forest (RF) adalah model yang menggunakan pohon keputusan untuk menangani data kompleks dengan berbagai fitur. Keunggulan utama RF adalah kemampuannya mengurangi *overfitting* melalui teknik ensemble learning [12]. *Support Vector Machine* (SVM) merupakan model berbasis margin yang berupaya menemukan *hyperplane* terbaik untuk membedakan kelas data. *Support Vector Machine* (SVM) memiliki kinerja yang baik dalam menangani jumlah data yang besar, terutama karena kemampuannya untuk memaksimalkan margin antar kelas, yang berkontribusi pada peningkatan generalisasi model [13]. *Logistic Regression* (LR) adalah metode yang sering digunakan dalam prediksi biner, seperti memprediksi keberadaan atau tidaknya penyakit gagal jantung. LR juga sering dijadikan acuan dalam penelitian klasifikasi medis karena sifatnya yang lebih mudah diinterpretasikan dibandingkan model lain [14].

Penelitian ini memanfaatkan dataset MIMIC-IV (*Medical Information Mart for Intensive Care IV*) yang didapatkan melalui situs *physionet* [15]. Data MIMIC-IV ini memiliki banyak sekali data pasien, salah satunya data pasien yang mengalami gagal jantung . Data MIMIC tersebut dapat diunduh setelah mendapatkan lisensi “*COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)*” [16]. Penggunaan data MIMIC-IV ini dilandaskan karena beberapa faktor. Faktor pertama, MIMIC-IV adalah salah satu sumber data rekam medis elektronik yang kini banyak dimanfaatkan dalam dunia penelitian medis dan pendidikan [9], [11]. Lalu faktor kedua dikarenakan data MIMIC-IV terbuka dan dapat diakses secara gratis, dataset ini memberikan peluang besar bagi siapa saja yang ingin melakukan studi tanpa terkendala oleh akses atau biaya [17]. Faktor

ketiga dikarenakan data yang tersedia sangat beragam dan mendalam, mencakup informasi penting seperti hasil laboratorium, riwayat diagnosa, prosedur medis, pemberian obat, hingga catatan dokter dalam bentuk teks naratif [17]. Dengan cakupan lebih dari 315.000 pasien, MIMIC-IV memungkinkan analisis yang kaya dan mendalam [17]. Namun, karena jumlah datanya sangat besar, diperlukan *preprocessing* data yang teliti, termasuk penanganan data yang hilang, normalisasi variabel, dan pemilihan fitur yang tepat agar model ML dapat bekerja dengan baik.

Penelitian ini bertujuan untuk membandingkan performa tiga model *machine learning* *Random Forest*, *Support Vector Machine*, dan *Logistic Regression* dalam memprediksi kejadian gagal jantung pada pasien ICU dengan menggunakan data MIMIC-IV. Pemilihan algoritma *Random Forest*, SVM, dan *Logistic Regression* dilakukan secara terarah berdasarkan penelitian terdahulu yang menunjukkan ketiganya sering digunakan untuk prediksi penyakit jantung dan memberikan hasil yang baik. Dengan mempertimbangkan relevansi klinis dan stabilitas model, ketiganya dipilih sebagai fokus utama dalam penelitian ini. Akurasi, ROC-AUC, *Recall*, dan melihat hasil dari *confusion matrix* untuk dijadikan acuan untuk menentukan model dengan performa terbaik dalam memprediksi gagal jantung di ICU. Keempat metrik ini digunakan karena saling melengkapi, akurasi mengukur proporsi prediksi yang benar [18], ROC-AUC mengevaluasi kemampuan model dalam membedakan antara pasien yang mengalami gagal jantung dan yang tidak [19], sedangkan *recall* (sensitivitas) mengukur kemampuan model untuk menemukan semua kasus positif yang relevan, yang sangat krusial dalam diagnosis medis di mana kegagalan mendeteksi penyakit dapat berakibat fatal [20]. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi faktor-faktor yang berpengaruh dalam prediksi gagal jantung, sehingga dapat membantu tenaga medis dalam mengambil keputusan yang lebih baik untuk menyelamatkan pasien gagal jantung di ICU.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disusun, berikut ini adalah rumusan masalah dalam penelitian ini:

1. Bagaimana performa model *Random Forest*, *Support Vector Machine* (SVM), dan *Logistic Regression* dalam memprediksi risiko gagal jantung pada pasien ICU?
2. Model machine learning mana (*Random Forest*, SVM, dan *Logistic Regression*) yang paling efektif dalam prediksi gagal jantung dengan menggunakan evaluasi akurasi, ROC-AUC, dan *Recall*?
3. Variabel apa saja yang paling berpengaruh dalam prediksi gagal jantung pada penelitian ini

1.3 Batasan Masalah

Dalam membandingkan kinerja algoritma *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Logistic Regression* (Logreg) untuk memprediksi kejadian gagal jantung pada pasien ICU menggunakan dataset MIMIC-IV, terdapat beberapa batasan masalah yang diterapkan. Berikut merupakan batasan masalahnya:

1. Fokus pada kesulitan deteksi gagal jantung di lingkungan ICU akibat gejala awal yang tidak spesifik dan perkembangan kondisi yang cepat. Meskipun dataset besar seperti MIMIC-IV tersedia, tantangan utama terletak pada identifikasi tanda-tanda awal yang samar dan dinamika kondisi pasien yang cepat berubah. Studi ini membatasi analisis pada faktor-faktor klinis dan laboratorium yang tersedia dalam dataset MIMIC-IV serta keterbatasan dalam interpretasi data waktu nyata (*real-time*).
2. Studi ini membandingkan kinerja tiga algoritma machine learning, yaitu *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Logistic Regression*, dalam memprediksi gagal jantung pada pasien ICU. Fokus utama adalah pada akurasi, ketahanan (*robustness*), dan kemampuan ketiga algoritma dalam menangani dataset medis yang kompleks seperti MIMIC-IV. Batasan mencakup penggunaan metrik evaluasi seperti akurasi, AUC-ROC, dan *Recall*.
3. Studi ini membatasi analisis pada kemampuan algoritma *Random Forest*, *Support Vector Machine*, dan *Logistic Regression*, dalam mengatasi tantangan dataset besar dan tidak seimbang (*imbalanced dataset*) seperti

MIMIC-IV. Fokusnya adalah pada teknik preprocessing data (seperti resampling, normalisasi, dan feature selection).

4. Hasil akhir dari penelitian ini berupa perbandingan performa model prediksi, bukan penerapan model dalam bentuk sistem atau aplikasi. Oleh karena itu, tahap deployment tidak dilakukan, karena fokus utama penelitian adalah untuk menganalisis dan membandingkan kinerja algoritma, bukan membangun sistem prediksi siap pakai.

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan tiga rumusan masalah pada bagian 1.3, penelitian ini menghasilkan tujuan yang ingin dicapai. Berikut adalah hasil dari penelitian ini:

1. Membandingkan kinerja algoritma machine learning, yaitu *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Logistic Regression*, dalam memprediksi penyakit jantung pada pasien ICU menggunakan dataset MIMIC-IV.
2. Mengidentifikasi variabel atau fitur paling signifikan dalam dataset MIMIC-IV yang berkontribusi terhadap prediksi penyakit jantung, agar dapat membantu tenaga medis dalam mengambil keputusan yang lebih baik untuk menyelamatkan pasien gagal jantung di ICU.
3. Mengembangkan model prediksi penyakit jantung yang dapat memberikan peringatan dini kepada tenaga medis untuk intervensi yang lebih cepat dan tepat.

1.4.2 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak. Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Penelitian ini memberikan wawasan tentang perbandingan kinerja algoritma *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Logistic Regression* dalam memprediksi kejadian penyakit

- jantung pada pasien. Hasil penelitian dapat menjadi referensi bagi pengembangan model prediksi yang lebih akurat di masa depan.
2. Dengan memanfaatkan dataset MIMIC-IV dan teknik preprocessing, penelitian ini bertujuan untuk meningkatkan akurasi serta ketahanan model dalam memprediksi penyakit jantung. Hal ini dapat membantu tenaga medis dalam mendeteksi dini kondisi kritis pasien.
 3. Penelitian ini menguji kemampuan algoritma *Random Forest*, *Support Vector Machine*, dan *Logistic Regression* dalam menangani dataset besar dan tidak seimbang, seperti MIMIC-IV. Hasilnya dapat menjadi acuan untuk pengembangan sistem prediksi berbasis machine learning pada dataset medis yang kompleks.

1.5 Sistematika Penulisan

Struktur penulisan penelitian ini terbagi ke dalam beberapa bab dan subbab. Adapun rincian sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada Bab I akan berisikan penjelasan penelitian ini. Penjelasan tersebut akan selaras dengan judul penelitian ini yaitu prediksi gagal jantung pasien ICU menggunakan algoritma *Random Forest*, *SVM*, dan *Logistic Regression*. Bab I membahas beberapa subbab, termasuk latar belakang yang berisikan latar belakang pengambilan judul tersebut sebagai penelitiannya, rumusan masalah yang berisikan masalah – masalah yang ingin dibenahi, tujuan penelitian berisikan tujuan dilakukannya penelitian ini serta bagaimana masalah yang telah dirumuskan akan diselesaikan, manfaat penelitian menjelaskan manfaat yang diperoleh dari hasil penelitian ini, sedangkan sistematika penulisan memaparkan struktur penyusunan laporan penelitian ini.

BAB II LANDASAN TEORI

Bab II secara spesifik membahas teori-teori yang mendukung penelitian ini. Teori yang dijelaskan dalam bab ini mencakup algoritma yang digunakan, yaitu SVM, *Random Forest*, dan *Logistic Regression*. CRISP-DM yang menjadi teknik data mining dalam penelitian ini, ROC-AUC, dan *recall* yang merupakan metrik evaluasi yang digunakan untuk mengukur kinerja model *Machine Learning* pada penelitian ini.

BAB III METODE PENELITIAN

Pada Bab III ini akan dijelaskan secara rinci objek penelitian serta metode penelitian yang digunakan dalam penelitian ini. Pada bab ini juga akan dijelaskan mengenai variabel penelitian, teknik pengumpulan data yang digunakan, teknik pengambilan sampel data, dan teknik analisis data yang digunakan dalam penelitian ini.

BAB IV HASIL DAN DISKUSI

Bab IV ini menyajikan hasil dari proses pengolahan data menggunakan algoritma SVM, *Logistic Regression*, dan *Random Forest* terhadap data pasien gagal jantung dari MIMIC-IV. Dibahas pula evaluasi model menggunakan metrik akurasi, ROC-AUC, dan *Recall*. Analisis dilakukan untuk mengetahui model terbaik dalam memprediksi gagal jantung pada pasien.

BAB V SIMPULAN DAN SARAN

Bab V ini memuat kesimpulan dari hasil penelitian yang telah dilakukan, termasuk model terbaik untuk prediksi gagal jantung pada pasien ICU berdasarkan data MIMIC-IV. Selain itu, disampaikan saran untuk penelitian selanjutnya agar hasil yang diperoleh lebih optimal.