

BAB I

PENDAHULUAN

1.1 Latar Belakang

Jantung adalah salah satu organ penting yang mempunyai fungsi untuk memompa darah ke seluruh tubuh. Apabila jantung mengalami kesulitan dalam memompa darah yang mengandung oksigen sesuai dengan kebutuhan tubuh, maka dapat meningkatkan kemungkinan terjadinya suatu kondisi medis yang disebut gagal jantung. Pada umumnya, gagal jantung dapat ditandai dengan beberapa gejala seperti kesulitan bernapas saat melakukan aktivitas fisik atau beristirahat, mudah lelah, dan mudah berkeringat di area perut, pergelangan tangan atau kaki. Beberapa faktor yang biasanya menyebabkan penyakit gagal jantung adalah diabetes, tekanan darah tinggi, anemia, pola hidup yang kurang sehat dan minimnya aktifitas fisik.

Menurut WHO (2020) penyakit jantung telah menjadi penyebab kematian tertinggi di dunia yang menyebabkan 16% dari total kematian di dunia. Sejak tahun 2000, menunjukkan peningkatan kematian dari 2 juta menjadi 8.9 juta kematian pada tahun 2019. Penyakit gagal jantung menjadi penyebab kematian tertinggi kedua di Indonesia setelah penyakit stroke. Berdasarkan diagnosis dokter yang dilaporkan oleh Riset Kesehatan Dasar Kementerian Kesehatan Indonesia pada tahun 2018, tercatat sekitar 1.5% atau sekitar 1.017.290 orang menderita gagal jantung di Indonesia. Penyakit gagal jantung menunjukkan peningkatan pasien dibandingkan dengan tahun 2013 yang hanya sekitar 0.5% atau sekitar 883.447 orang (Rokom, 2022). Hal ini yang menjadi masalah serius dikarenakan pentingnya organ vital jantung dan tingginya tingkat kematian yang diakibatkan penyakit gagal jantung. Dengan tingkat kematian yang tinggi, maka diperlukan perhatian yang lebih untuk mendeteksi awal pasien gagal jantung.

Perkembangan teknologi telah memberikan kontribusi dalam mengatasi permasalahan di berbagai bidang terutama pada bidang kesehatan, salah satunya dalam pengklasifikasian dan deteksi penyakit menggunakan metode *machine learning*. *Machine learning* atau pembelajaran mesin merupakan suatu cabang dari kecerdasan buatan yang mempelajari cara komputer belajar dari data dengan tujuan meningkatkan kecerdasannya (Wahyono, 2018). Beberapa metode yang

umum digunakan dalam *machine learning* adalah *Artificial Neural Network*, *K-Nearest Neighbors*, *Decision Tree*, *Random Forest*, *Naive Bayes*, dan *Support Vector Machine*. Dalam mendeteksi awal kelangsungan hidup pasien gagal jantung, penelitian ini akan menggunakan metode *Random Forest*. Hal ini dikarenakan metode *Random Forest* dapat memberikan tingkat akurasi yang tinggi dan dapat menangani data yang besar dan kompleks. Metode *Random Forest* adalah metode *ensemble* yang menggabungkan hasil dari beberapa pohon keputusan yang dapat meningkatkan akurasi dan mengurangi masalah *overfitting* serta dapat mengidentifikasi fitur yang paling penting untuk digunakan pada data latih (Polamuri, 2017). Meskipun *Random Forest* memiliki kemampuan untuk mengatasi data dalam jumlah besar, bukan berarti mendefinisikan model yang optimal bisa dilakukan dengan mudah. Untuk menghasilkan model yang optimal perlu dilakukan evaluasi model secara berkala dan memperhatikan faktor-faktor lain seperti jumlah sampel, kualitas data, dan menentukan parameter (Azhar et al., 2021).

Dalam membangun model *Random Forest*, tantangan yang sering muncul adalah menentukan parameter yang optimal untuk meningkatkan kinerja model. Menentukan kombinasi parameter yang tepat secara manual bisa menjadi rumit dan menghabiskan waktu. Maka, pilihan tepat yang dapat dilakukan untuk mendapatkan kombinasi parameter terbaik yaitu *hyperparameter tuning* (Shekar & Dagneu, 2019). Dalam proses *hyperparameter tuning*, terdapat dua metode yang umum digunakan adalah *Grid Search* dan *Random Search*. Kedua metode tersebut bertujuan untuk mencari kombinasi hyperparameter yang memberikan kinerja model yang optimal, tetapi keduanya mengadopsi pendekatan yang berbeda dalam mengeksplorasi ruang hyperparameter.

Penggunaan *machine learning* dengan metode *Random Forest* dalam deteksi kelangsungan hidup merujuk pada kategorisasi pasien menjadi dua kelompok yaitu selamat dan tidak selamat. Status selamat yaitu pasien yang masih hidup setelah mengalami gagal jantung, sedangkan tidak selamat yaitu pasien yang telah meninggal dunia. Informasi mengenai status selamat dan tidak selamat ini memberikan gambaran tentang hasil dari kondisi kesehatan pasien setelah dilakukan tindakan lanjut awal. Jadi, prediksi ini mengidentifikasi apakah pasien memiliki peluang yang lebih tinggi untuk selamat atau tidak selamat dari kondisi gagal jantung tanpa memberikan estimasi waktu. Faktor yang digunakan dalam penelitian ini adalah usia, anemia, hipertensi, enzim cpk, diabetes, fraksi ejeksi, trombosit, jenis kelamin, serum kreatinin, serum sodium, merokok, dan waktu tindak lanjut.

Penelitian terdahulu mengenai kelangsungan hidup pasien penyakit gagal jantung telah dilakukan oleh Ahmad,dkk (2017) dengan judul penelitian “*Survival analysis of heart failure patients: a case study*”, penelitian ini menggunakan algoritma Regresi Cox yaitu teknik regresi yang digunakan untuk menganalisis faktor-faktor yang memengaruhi waktu kelangsungan hidup. Hasil dari penelitian ini menunjukkan bahwa dengan faktor-faktor seperti usia, disfungsi ginjal, tekanan darah tinggi, fraksi ejeksi rendah, dan anemia secara signifikan meningkatkan risiko kematian pada pasien gagal jantung. Penelitian serupa mengenai kelangsungan hidup pasien gagal jantung telah dilakukan oleh Prahasti dan Fauzi (2021) dengan judul penelitian “Risiko Kematian Pasien Gagal Jantung Kongestif (GJK): Studi kohort Retrospektif berbasis Rumah Sakit”. Hasil dari penelitian terdapat 3 variabel yang secara signifikan mempengaruhi risiko kematian pada gagal jantung bahkan setelah mempertimbangkan variabel lain, yaitu status anemia, denyut jantung dan status trombositopenia.

Penelitian terdahulu selanjutnya dilakukan dengan membandingkan algoritma *Random Forest* dengan algoritma *Artificial Neural Network* dalam diagnosa penyakit yang telah dilakukan oleh Eky Cahya Putra Witjaksana, dkk (2021) dengan judul “Perbandingan akurasi algoritma *Random Forest* dan algoritma *Artificial Neural Network* untuk klasifikasi penyakit diabetes”. Penelitian ini dilakukan untuk mengetahui algoritma mana yang lebih bagus dalam klasifikasi penyakit diabetes. Hasil dari penelitian ini adalah algoritma *Random Forest* memiliki akurasi 90.62% sedangkan algoritma *Artificial Neural Network* memiliki akurasi 82.29%. Hal ini menunjukkan bahwa algoritma *Random Forest* lebih baik dibandingkan algoritma *Artificial Neural Network* dalam mengklasifikasikan penyakit diabetes.

Penelitian lainnya dilakukan dengan membandingkan algoritma *Random Forest* dengan *Support Vector Machine* dalam diagnosa penyakit yang telah dilakukan oleh Chea Zahra Vaganza Junus, dkk (2022) dengan judul “Klasifikasi menggunakan metode *Support Vector Machine* dan *Random Forest* untuk deteksi awal risiko diabetes melitus”. Penelitian ini menerapkan *hyperparameter tuning* pada kedua algoritma dengan tujuan mendapatkan parameter terbaik dari setiap algoritma. Hasil dari penelitian ini menunjukkan bahwa algoritma *Random Forest* dengan *hyperparameter tuning* menghasilkan kinerja klasifikasi yang baik dengan akurasi 98.08% dibandingkan algoritma *Support Vector Machine* dengan akurasi 91%.

Berdasarkan latar belakang yang telah diuraikan, maka penulis melakukan penelitian dengan judul penelitian “Deteksi Awal Kelangsungan Hidup Pasien Gagal Jantung Menggunakan *Machine Learning* Metode *Random Forest*”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka identifikasi masalah dalam penulisan penelitian ini adalah sebagai berikut:

1. Penyakit gagal jantung menjadi penyebab kematian tertinggi di dunia dan kedua di Indonesia.
2. Dalam *machine learning* menggunakan metode *Random Forest* memiliki parameter yang perlu diatur dengan baik agar menghasilkan kinerja yang optimal. Namun, menentukan kombinasi parameter yang tepat secara manual bisa menjadi rumit dan memakan waktu.
3. Metode *Random Forest* sangat bergantung pada pengaturan parameter yang optimal. *Hyperparameter tuning* adalah langkah penting dalam meningkatkan kinerja model.

1.3 Ruang Lingkup

Ruang lingkup bahasan pada penelitian ini terfokus pada penerapan *machine learning* metode *Random Forest* dan peningkatan kinerja model lebih optimal menggunakan *hyperparameter tuning* untuk mendeteksi awal kelangsungan hidup pasien dengan kondisi gagal jantung.

1.4 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka rumusan masalah dalam penulisan penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja *Random Forest* dalam mendeteksi awal kelangsungan hidup pasien gagal jantung sebelum menggunakan *hyperparameter tuning* berdasarkan metrik evaluasi model?
2. Bagaimana hasil peningkatan kinerja *Random Forest* menggunakan *hyperparameter tuning* dalam mendeteksi awal kelangsungan hidup pasien gagal jantung berdasarkan metrik evaluasi model?

1.5 Batasan Masalah

Dalam penelitian ini diperlukan batasan masalah untuk menghindari meluasnya permasalahan yang akan dibahas yaitu:

1. Data yang digunakan dalam penelitian ini diperoleh dari situs UCI *Machine Learning Repository* tahun 2020 yang berjudul “*Heart Failure Clinical Records Dataset*”. Data yang diperoleh merupakan data sekunder yang berupa data rekam medis 299 pasien yang mengalami gagal jantung dan memiliki 12 variabel prediktor dan 1 variabel target.
2. Penelitian ini menggunakan metode *Random Forest* untuk membangun model dan *hyperparameter tuning* untuk meningkatkan kinerja model.
3. Prediksi kelangsungan hidup merujuk pada kategorisasi pasien menjadi dua kelompok yaitu selamat dan tidak selamat berdasarkan informasi yang ada dalam dataset klinis.

1.6 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penulisan penelitian ini adalah sebagai berikut:

1. Mengetahui kinerja *Random Forest* dalam mendeteksi awal kelangsungan hidup pasien gagal jantung sebelum menggunakan *hyperparameter tuning* berdasarkan metrik evaluasi model.
2. Mengetahui hasil peningkatan kinerja *Random Forest* menggunakan *hyperparameter tuning* dalam mendeteksi awal kelangsungan hidup pasien gagal jantung berdasarkan metrik evaluasi model.

1.7 Manfaat Penelitian

Dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi penulis, penelitian ini berguna untuk menambah pengetahuan dan pemahaman dalam menggunakan metode *Random Forest*.
2. Bagi pembaca, penelitian ini berguna untuk menambah pengetahuan dan dapat menjadi referensi tambahan oleh praktik klinis untuk mendukung deteksi awal kelangsungan hidup pasien gagal jantung.
3. Bagi akademik, penelitian ini dapat dipergunakan sebagai tambahan informasi dan sumber bacaan mengenai metode *Random Forest* bagi yang hendak melakukan penelitian serupa dan penelitian selanjutnya.