

BAB I PENDAHULUAN

1.1 Latar Belakang

Penyakit *kardiovaskular* merupakan penyebab kematian utama di seluruh dunia dengan estimasi 19,8 juta kematian pada 2022, mewakili 32% dari total kematian global [1]. Serangan jantung atau penyakit jantung *iskemik* menunjukkan tren peningkatan beban penyakit yang signifikan, dengan jumlah kematian akibat *CVD* meningkat dari 13,1 juta pada tahun 1990 menjadi 19,2 juta pada tahun 2023 [2]. Di Indonesia, prevalensi penyakit jantung berdasarkan diagnosis dokter pada tahun 2023 menunjukkan gradien yang signifikan berdasarkan kelompok usia [3]. Sistem kesehatan Indonesia menghadapi keterbatasan tenaga spesialis jantung dengan hanya 1.708 kardiolog dari target ideal 2.700 kardiolog, serta distribusi yang tidak merata di berbagai wilayah [4].

Perkembangan *machine learning* dan *deep learning* telah merevolusi diagnosis medis dengan meningkatkan akurasi dan efisiensi dalam deteksi, dan klasifikasi [5]. Abbas et al. mengembangkan model *deep learning* untuk deteksi *infark miokard* dengan akurasi 94,3% [6], sementara Waleed et al. membuktikan efektivitas *ensemble learning* dengan akurasi 98,7% [7]. Namun, mayoritas model *machine learning* memiliki karakteristik "*black box*" yang menyulitkan interpretasi proses pengambilan keputusan [8]. Dalam konteks medis, ketidaktransparanan ini menjadi hambatan serius bagi adopsi teknologi *AI* karena praktisi kesehatan memerlukan pemahaman yang jelas mengenai alasan di balik setiap prediksi untuk membangun kepercayaan dan memastikan keamanan pasien [9].

Studi terdahulu menunjukkan bahwa pendekatan *Explainable AI (XAI)* dapat digunakan untuk mengatasi masalah interpretabilitas pada model prediksi medis [10]. Park et al. menggunakan teknik *Local Interpretable Model-agnostic Explanations (LIME)* untuk memberikan penjelasan pada model prediksi serangan jantung berbasis *ensemble learning* [11]. Meskipun demikian, metode *LIME* memiliki keterbatasan mendasar karena mengandalkan proses perturbasi acak, yang

dapat mengakibatkan ketidakstabilan dalam penjelasan dan menurunkan keandalan interpretasi [12]. Di sisi lain, algoritma *Logistic Regression* memiliki keunggulan signifikan dalam hal interpretabilitas [13]. Setiap fitur memiliki koefisien yang dapat diinterpretasikan secara langsung sebagai kontribusi terhadap probabilitas *output* [14]. Meskipun tergolong sederhana, model *Logistic Regression* dapat mencapai peningkatan performa yang signifikan melalui proses optimasi *hyperparameter* [15]. Penelitian pada prediksi penyakit jantung menunjukkan bahwa model yang dioptimasi dapat mencapai akurasi hingga 90,16% [16]. Untuk meningkatkan interpretabilitas lebih lanjut, *SHapley Additive exPlanations (SHAP)* berbasis teori permainan menawarkan pendekatan yang konsisten pada level global maupun lokal [17]. Selain itu, analisis empiris terbaru menunjukkan bahwa *SHAP* terbukti memiliki *robustness* yang kuat, mampu mempertahankan kestabilan peringkat pentingnya fitur meskipun model beroperasi dalam kondisi data yang tidak sempurna atau terkontaminasi [18]. Kombinasi antara *Tuned Logistic Regression* dengan *SHAP* berpotensi menghasilkan model yang tidak hanya akurat tetapi juga dapat menjelaskan keputusan prediksi secara transparan, namun pendekatan ini belum banyak dieksplorasi dalam konteks deteksi dini serangan jantung.

Berdasarkan kondisi tersebut, penelitian ini mengembangkan model *Tuned Logistic Regression* dan diintegrasikan dengan *Explainable AI* menggunakan pendekatan *SHAP* untuk deteksi serangan jantung. Pendekatan ini diharapkan dapat menjembatani *trade-off* antara akurasi dan interpretabilitas, menghasilkan model yang tidak hanya mampu mendeteksi serangan jantung dengan akurat tetapi juga dapat menjelaskan dasar pengambilan keputusan secara transparan kepada praktisi medis.

1.2 Rumusan Masalah

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

1. Seberapa besar peningkatan kinerja model *Logistic Regression* yang dioptimasi dengan *hyperparameter tuning* dibandingkan *baseline* model dalam deteksi serangan jantung?
2. Sejauh mana *robustness* dan interpretabilitas yang dihasilkan *SHAP* dalam menjelaskan kontribusi fitur-fitur klinis terhadap prediksi model *Tuned Logistic Regression* untuk deteksi dini serangan jantung?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan terfokus, maka ditetapkan batasan masalah sebagai berikut:

1. Model yang dikembangkan terbatas pada *Logistic Regression* dengan optimasi *hyperparameter* menggunakan *GridSearchCV*, *Randomized Search*, dan *Bayesian Optimization*.
2. Dataset yang digunakan adalah dataset publik penyakit jantung dalam bentuk data tabular berisi fitur klinis pasien, dengan fokus hanya pada variabel yang tersedia di dataset.
3. Metode interpretabilitas terbatas pada *SHAP* (*SHapley Additive exPlanations*) tanpa melibatkan metode *Explainable AI* lainnya.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitian ini adalah:

1. Mengoptimasi model *Logistic Regression* dengan *hyperparameter tuning* dan mengukur peningkatan kinerjanya terhadap *baseline* model menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *AUC-ROC*.
2. Mengimplementasikan dan menganalisis *robustness* serta interpretabilitas *SHAP* dalam menjelaskan kontribusi fitur klinis pada prediksi model *Tuned Logistic Regression*.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Secara Teoritis

Penelitian ini berkontribusi pada pengembangan ilmu *data science* dan kesehatan dengan memperkaya literatur mengenai penerapan *Tuned Logistic Regression* yang diintegrasikan dengan *SHAP* untuk deteksi dini serangan jantung.

2. Secara Praktis

Penelitian ini dapat membantu *user* deteksi dini melalui model yang lebih transparan sebagai sistem pendukung keputusan, serta menjadi referensi bagi peneliti selanjutnya dalam mengembangkan model *Explainable AI* pada kasus medis lain.

1.6 Sistematika Penulisan

Untuk memberikan gambaran yang jelas serta menjaga fokus pembahasan, maka skripsi ini disusun dengan sistematika penulisan sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian teori terkait penyakit jantung, metode *Logistic Regression*, optimasi parameter (*tuning*), *Explainable AI* dengan *SHAP*, serta penelitian terdahulu yang relevan.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan metode penelitian yang digunakan, meliputi alur penelitian, pengumpulan dan pengolahan data, metode yang diusulkan, serta teknik evaluasi model.

4. BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil implementasi model *Tuned Logistic Regression* dengan *SHAP* untuk deteksi serangan jantung, evaluasi performa model,

visualisasi interpretasi, serta pembahasan hasil penelitian.

5. BAB V PENUTUP

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan serta saran untuk penelitian selanjutnya.

