

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Jantung adalah organ yang esensial dalam tubuh manusia dan berfungsi sebagai pusat pengatur dalam sistem sirkulasi, dimana pembuluh darah bertindak sebagai jalur distribusinya [1]. Peranan jantung sangat krusial dalam memasok oksigen ke seluruh tubuh dan mengeliminasi hasil metabolisme yang tidak dibutuhkan [2]. Meskipun memegang peran vital, jantung juga merupakan organ yang sangat rentan terhadap berbagai penyakit.

Menurut data dari *Global Burden of Disease* dan *Institute for Health Metrics and Evaluation* (IHME) pada periode 2014-2019, penyakit jantung adalah penyebab utama kematian di Indonesia. Data Riset Kesehatan Dasar (Riskedas) mencatat adanya tren peningkatan angka kejadian penyakit jantung dari 0,5% pada tahun 2013 menjadi 1,5% pada tahun 2018 [3]. Menurut informasi dari WHO dan CDC sekitar 11,3% dari populasi dewasa di negara-negara berkembang atau sekitar 26,6 juta orang menderita penyakit jantung [4]. Di Indonesia, jumlah orang yang menderita penyakit jantung pada tahun 2020 meningkat sekitar 24% dibandingkan dengan tahun 2005, yang setara dengan sekitar 82 juta jiwa yang terkena penyakit tersebut [5]. Pada tahun 2022, Indonesia mengalami lonjakan kasus penyakit serius, termasuk penyakit jantung dengan jumlah kasus mencapai 15,5 juta, kanker sebanyak 3,2 juta kasus, stroke sekitar 2,5 juta kasus, dan gagal ginjal dengan jumlah kasus mencapai 1,3 juta [6].

Namun sebagian besar penyakit ini disebabkan oleh gaya hidup yang buruk seperti pola makan tidak sehat, kurangnya olahraga, merokok, minum alkohol, dan stres tinggi. Penyakit ini bisa dicegah dengan deteksi dini dan pengendalian risiko. Penting untuk mendeteksi penyakit jantung secara awal, terutama jika ada faktor bawaan, untuk mencegah risiko komplikasi serius dan kematian yang dapat dihindari [7]. Ketidaktahuan tentang gejala awal penyakit jantung dan kurangnya praktik gaya hidup sehat menyebabkan masyarakat tidak menyadari risikonya. Saat ini, deteksi penyakit jantung masih dilakukan secara manual dengan berkonsultasi ke rumah sakit dan dokter, diikuti dengan pemeriksaan laboratorium dan tindak lanjut dokter. Penting untuk menjalani pemeriksaan secara rutin guna mendeteksi penyakit jantung sejak dini untuk pencegahan atau penanganan yang tepat. Ini memerlukan biaya yang signifikan. Oleh karena itu, dibutuhkan sistem deteksi penyakit jantung yang dapat memberikan hasil akurat dan lebih efisien secara finansial bagi masyarakat [8][9].

Dengan perkembangan teknologi, angka kematian akibat penyakit jantung dapat dikurangi. Machine learning (ML) bisa menjadi solusi untuk membantu profesional kesehatan dalam mengambil keputusan yang tepat terkait diagnosis pasien [10]. Machine learning adalah metode yang mengubah data mentah menjadi informasi yang membantu dalam pengklasifikasian dan prediksi [11]. Meskipun tidak menjamin keakuratan 100%, Machine Learning seringkali efektif dalam mengidentifikasi penyakit jantung, membuktikan kebermanfaatannya dalam bidang kesehatan [12]. Beberapa penelitian sebelumnya telah dilakukan mengenai klasifikasi penyakit jantung menggunakan *machine learning*. Salah satunya penelitian Hidayat dkk [13] melakukan klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier yang mendapatkan akurasi sebesar 94% yang dimana memiliki 2 label yaitu 0 (normal) dan 1(penyakit jantung), Sedangkan penelitian yang dilakukan oleh Arni Sepharni dkk [14] mengusulkan algoritma C4.5 dalam klasifikasi penyakit jantung mendapatkan akurasi sebesar 79%. Lalu penelitian yang selanjutnya dilakukan oleh Dewi Nasien dkk [15] melakukan klasifikasi penyakit jantung menggunakan Decision Tree dan KNN serta Ekstraksi Fitur PCA mendapatkan akurasi KNN yang lebih baik yang mendapatkan akurasi sebesar 81.82%, precision 87.04%, recall 79.82%, dan F1 score 83.13%.

Berdasarkan paparan penelien diatas, implementasi algoritma klasifikasi yang merupakan pendekatan *machine learning* terbukti dapat memberikan akurasi yang baik untuk mengklasifikasi penyakit jantung, sehingga penelitian yang dilakukan akan mengimplementasikan perbandingan beberapa algoritma klasifikasi untuk mengklasifikasi penyakit jantung yang menggunakan data terbaru yang didapatkan dari sumber terbuka yaitu kaggle. Sehingga dapat disimpulkan penelitian yang akan dilakukan mengangkat topik penelitian dengan judul **“Perbandingan Algoritma Klasifikasi Pada Pasien Penyakit Jantung”**.

1.2 Rumusan Masalah

Perumusan masalah penelitian yang dilakukan adalah melakukan perbandingan algoritma klasifikasi pada pasien penyakit jantung

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Tujuan dari penelitian yang dilakukan adalah untuk mengimplementasikan algoritma klasifikasi pada pasien penyakit jantung, Setelah itu membandingkan algoritma tersebut berdasarkan akurasi sehingga didapatkan algoritma yang lebih optimal terhadap pasien penyakit jantung.

1.3.2 Manfaat

Adapun beberapa manfaat yang didapatkan dari penelitian yang telah dilakukan diantara lain sebagai berikut

1. Membantu praktisi medis dalam mengklasifikasikan pasien penyakit jantung dan normal dengan menggunakan data data yang dimasukkan sebagai referensi
2. Membantu dalam personalisasi perawatan untuk setiap pasien berdasarkan profil risiko individu sehingga meningkatkan efektivitas intervensi medis
3. Membantu klasifikasi untuk menganalisis data pasien, sehingga dapat membantu dalam prediksi dini penyakit jantung pada individu yang berisiko tinggi.

1.4 Batasan Masalah

1. Dataset yang digunakan berbentuk tabular atau csv yang diambil dari <https://www.kaggle.com/datasets/mexwell/heart-disease-dataset/data>
2. Metode yang diterapkan menggunakan algoritma klasifikasi
3. Model yang di implementasikan adalah RandomForestClassifier, CatBoostClassifier, HistGradientBoostingClassifier, XGBClassifier, VotingClassifier
4. Bahasa pemrograman yang diterapkan adalah Python 3
5. Parameter yang diterapkan sebagai ukuran adalah Normal dan Penyakit Jantung

1.5 Keterbaruan

1. Penelitian yang dilakukan oleh Jalis Dwi Muthohhar dkk [16] membandingkan algoritma klasifikasi untuk penyakit jantung yang menggunakan model Random Forest Classifier, Decision Tree, dan Naïve Bayes, data yang digunakan *Heart Attack Analysis & Prediction Dataset* yang diambil dari kaggle. Penelitian tersebut mendapatkan hasil pengujian model *decision tree* menggunakan evaluasi *random search* mendapatkan akurasi 84%, kemudian model *naïve bayes* 85% menggunakan *random forest* dan *grid*

search, Lalu untuk model *random forest classifier* menggunakan evaluasi *random search* mendapatkan akurasi 85% dan *grid search* 86%.

2. Penelitian yang dilakukan oleh Abram Setyo Prabowo [17] melakukan analisis perbandingan kinerja algoritma klasifikasi dalam deteksi penyakit jantung dengan model *Support Vector Machine (SVM)*, *Random Forest*, *Logistic Regression*, dan *AdaBoost*, yang dimana hasil pengujian dari ketiga model tersebut yang mendapatkan akurasi tertinggi adalah *AdaBoost* dengan hasil pengujian 98% akurasi, 98% recall, dan f1-score 98%.
3. Penelitian yang dilakukan oleh Made Krisna Dwipa Jaya dkk [18] melakukan perbandingan terhadap 4 algoritma klasifikasi yaitu *Random Forest*, *Decision Tree*, *Gradient Boosting*, *Logistic Regression* untuk klasifikasi penyakit jantung yang dapat disimpulkan dari 4 algoritma tersebut yang terbaik adalah *Random Forest* mendapatkan nilai *recall* sebesar 80,6% dan ROC AUC sebesar 76,3%.
4. Penelitian yang dilakukan oleh Anis Fitri Nur Masruriyah dkk [19] melakukan pengukuran kinerja model klasifikasi dengan data *oversampling* pada algoritma *supervised learning* untuk penyakit jantung yang dimana penelitian tersebut menggunakan teknik *oversampling* seperti SMOTE dan ADASYN untuk meningkatkan kemampuan membedakan kelas penyakit jantung dan normal, model terbaik menggunakan *random forest* dengan SMOTE mencapai nilai AUC sebesar 86%
5. Penelitian yang dilakukan oleh Ratnasari dkk [20] menggunakan *machine learning* untuk klasifikasi penyakit jantung, Dalam penelitian tersebut membandingkan enam algoritma klasifikasi machine learning dengan dataset dari *Cleveland Clinic Foundation*. *Naïve Bayes* memiliki akurasi tertinggi 84,67%, diikuti oleh *Logistic Regression* 84,30%, *Random Forest* mencapai 81,70%, *Decision Tree* 74%, KNN 73% dan *AdaBoost* 71,30%.
6. Penelitian yang dilakukan oleh Aldi Fianda Putra dkk [21] melakukan klasifikasi serangan jantung miokarditis yang *imbalance* dengan algoritma *Decision Tree* dan *Smote* yang dimana hasil eksperimen menunjukkan bahwa algoritma *Decision Tree* memiliki akurasi lebih tinggi daripada *Naïve Bayes* dan SVM, baik sebelum maupun setelah *oversampling*, dengan akurasi konsisten di atas 95\$. Sehingga, untuk dataset campuran kategoris dan numerik