

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Jantung manusia berperan sebagai organ vital yang memompa darah ke seluruh tubuh melalui empat ruang, yakni dua atrium dan dua ventrikel[4]. Jantung juga memiliki peranan untuk melakukan koordinasi dengan organ penting lain seperti jantung, otak, lambung dan ginjal[5]. Dengan begitu penyakit pada jantung akan sangat beresiko tinggi bagi penderitanya. Penyakit jantung ini akan menyebabkan terjadinya gagal jantung (*Heart Failure*) sehingga dapat menyebabkan kematian.[6]. Penyakit jantung yang terjadi dapat berupa disfungsi ventrikel kanan, kelainan katup jantung dan penyakit jantung bawaan. Penyakit Jantung disebabkan oleh pola hidup tidak sehat dan obesitas[7].

Berdasarkan data WHO (World Health Organization), penyakit jantung merupakan penyebab utama kematian di dunia, yaitu sebesar 32%. Oleh karena itu, penting untuk dilakukan pendeteksian dini terhadap penderita penyakit jantung. Pendeteksian dini pada penyakit jantung terdapat beberapa metode medis seperti Elektrokardiogram dan *stress test* pada jantung [8, 9]. Pada saat ini, terdapat metode *Angiography* yang banyak digunakan pihak medis yang dipercaya tepat untuk mendeteksi penyakit jantung. Namun metode ini terdapat efek samping terhadap pasien dan harga yang lebih tinggi, sehingga terjadi limitasi dalam pendeteksian terhadap penyakit jantung. Adapun metode yang dilakukan secara konvensional yang diterapkan dengan melihat riwayat pasien dan keluhan yang dilaporkan, hal ini membuat proses diagnosa lebih rumit dan bisa menimbulkan kesalahan diagnosa dari manusia. Dari permasalahan tersebut dapat dilakukan pendeteksian dini menggunakan aplikasi yang mampu melakukan deteksi penyakit jantung berdasarkan kondisi penderita [10, 11]. Pendeteksian dini ditujukan untuk mengambil keputusan dalam tahap awal untuk mendeteksi penyakit jantung tersebut. Dengan adanya hasil dari deteksi dini dapat dilakukan penanganan secara konseling maupun dalam bentuk pengobatan secara dini [7]. Deteksi dini dapat dilakukan menggunakan *machine learning* [10, 11, 12]. *Machine learning* telah digunakan untuk memahami kompleksitas dan interaksi non-linier antara berbagai faktor dengan mengurangi kesalahan dalam prediksi dan hasil faktual [13]. Penggunaan *machine learning* pada deteksi dini penyakit jantung bisa membantu

dalam mengambil keputusan dan mengurangi cost untuk pemeriksaan lainnya [11].

Ensemble Learning adalah pendekatan dalam pembelajaran mesin di mana berbagai model, seperti *base learners* atau *classifiers*, disatukan secara strategis untuk meningkatkan akurasi prediksi dan kinerja generalisasi. Metode ensemble meningkatkan kinerja dengan menggabungkan *output* dari *multiple model* untuk mengurangi varians keseluruhan sehingga dapat meningkatkan akurasi dan mengurangi risiko *overfitting* [14, 15]. Pada *ensemble learning* terdapat salah satu teknik yang dapat digunakan adalah teknik *boosting*[16].

Boosting merupakan teknik dalam *ensemble learning* yang dapat menguasai tugas-tugas kompleks dengan menggabungkan kesalahan dari model-model yang digunakan secara berurutan untuk meningkatkan kinerja prediksi atau klasifikasi secara bertahap [17]. Teknik *boosting* ini telah digunakan di beberapa penelitian seperti klasifikasi konsentrasi PM2.5 pada udara [18] dan pendeteksian penyakit jantung [19]. Pada teknik *Boosting* memiliki berbagai macam algoritma, salah satu algoritma tersebut adalah *Light Gradient Boosting Machine (LightGBM)*.

Light Gradient Boosting Machine (LightGBM) merupakan kerangka kerja peningkatan gradien yang menggunakan algoritma pembelajaran berbasis pohon[20]. Algoritma ini melakukan komputasi *parallel* seperti *feature parallel*, *data parallel* dan *voting parallel* yang bertujuan untuk meningkatkan efisiensi komputasi[21, 2]. Keunggulan pada algoritma LightGBM yaitu memiliki teknik *Gradient-based One-Side Sampling (GOSS)* dan *Exclusive Feature Bundling (EFB)*. Teknik *Gradient-based One-Side Sampling* membuat algoritma ini mampu mempelajari lagi dari hasil prediksi yang salah sebelumnya sehingga memungkinkan algoritma memperoleh hasil yang lebih baik. Teknik *Exclusive Feature Bundling* ini memungkinkan untuk menggabungkan banyak fitur eksklusif ke dalam sebuah kelompok fitur yang lebih padat sehingga membuat waktu komputasi menjadi lebih cepat[18]. Algoritma LightGBM telah digunakan pada beberapa penelitian, seperti deteksi *cyberattack* dari berbagai *network* [22], klasifikasi kondisi abnormal pada pengemudi sepeda motor [23] dan klasifikasi website *phising* [24].

Tabel 1.1. Tabel perbandingan penelitian terkait dalam prediksi penyakit jantung

Judul	Algoritma	Akurasi Terbaik	Feature yang digunakan	Hyperparameter Tuning
Heart failure survival prediction using machine learning algorithm: Am i safe from heart failure?	Logistic Regression, Support Vector Machine, XGBoost, LightGBM, Decision Tree, dan Bagging	85%	Age, Anemia, High Blood Pressure, Creatinine phosphokinase (CPK), Diabetes, Ejection fraction, Platelets, Sex, Serum creatinine, Serum sodium, Smoking, Time	learning_rate=0.1, reg_alpha=0.0, reg_lambda=0.0, number_leaves=31, max_depth=-1, n_estimators=100
Predicting coronary heart disease using an improved lightgbm model: Performance analysis and comparison	AdaBoost, DecisionTree, Bagging, XGBoost, CatBoost, dan LightGBM	93%	Sex, Age, Education, CurrentSmoker, CigsPerDay, BPMeds, PrevalentStroke, PrevalentHyp, Diabetes, TotChol, SysBp, DiaBp, BMI, Heart Rate, Glucose, TenYearCHD	learning_rate=0.0246, reg_alpha=0.0011, reg_lambda=0.0131, number_leaves=346, max_depth=-117, n_estimators=726
Heart disease risk assessment by using lightgbm technique	LightGBM	83.67%	num_major_vessels, chest_pain_type, st_slope, max_heart_rate_achieved, thalassemia, exercise_induced_angina, rest_electrocardiographic, fasting_blood_sugar, resting_blood_pressure, sex, st_depression, cholesterol, age	learning_rate=0.1, reg_alpha=0.0, reg_lambda=0.0, number_leaves=31, max_depth=-1, n_estimators=100

Penelitian serupa terkait penyakit jantung yang melakukan perbandingan berbagai algoritma[25]. Algoritma yang dibandingkan berupa Logistic Regression, Support Vector Machine, XGBoost, LightGBM, Decision Tree, dan Bagging dengan menggunakan dataset *Heart Disease UCI*. Dataset yang digunakan terdiri dari 299 data dan 13 *features*. Hasil dari penelitian tersebut adalah algoritma LightGBM memiliki akurasi tertinggi sebesar 85%, *precision* sebesar 87.27%, *recall* sebesar 82.64%, *f1-score* sebesar 84.55%, dan *Area Under Curve* (AUC) sebesar 93%.

Penelitian serupa lainnya terkait deteksi penyakit jantung yang

menggunakan dataset sebanyak 4240 data dan 15 *features* [20]. Penelitian tersebut melakukan perbandingan performa antara beberapa teknik dan algoritma klasifikasi seperti AdaBoost, DecisionTree, Bagging, XGBoost, CatBoost, dan LightGBM menggunakan dataset CHD (*Coronary Heart Disease*) dari *Framingham Heart Institute*. Hasil yang didapatkan adalah algoritma LightGBM berhasil mencapai tingkat akurasi tertinggi sebesar 92.4%, *precision* sebesar 96.2%, *recall* sebesar 88.7%, *f1-score* sebesar 0.923 dan *AUC-ROC* sebesar 93.2%. Hasil tersebut didapatkan berdasarkan *hyperparameter tuning* yang meliputi *max_depth*=117, *lambda_l1*=0.00115, *lambda_l2*=0.0131, *num_leaves*=346, *learning_rate*=0.0219, *n_estimators*=726, *feature_fraction*=0.5066, *bagging_fraction*=0.56, *bagging_freq*=7 and *min_child_samples*=2.

Penelitian serupa lainnya terkait deteksi penyakit jantung menggunakan algoritma LightGBM [12]. Dataset yang digunakan memiliki 4990 data dan 28 *features* yang berasal dari dataset *heart disease* Cleveland UCI. Dalam percobaan tersebut, algoritma LightGBM mampu menunjukkan tingkat akurasi sebesar 83.67%, *precision* sebesar 97.5%, *recall* sebesar 97.5%, *f1-score* sebesar 0.975, *Area Under Curve* (AUC) sebesar 96.7% dan *Receiver operating characteristic* (ROC) sebesar 96.7%.

Berdasarkan latar belakang yang telah dijelaskan, penelitian ini akan menerapkan algoritma LightGBM dalam melakukan deteksi penyakit jantung berdasarkan dataset *heart disease UCI Machine Learning Repository*. Lalu berdasarkan penelitian terkait sebelumnya dilakukan pengembangan dari feature yang digunakan dan dilakukan *hyperparameter tuning*.

1.2 Rumusan Masalah

Dari latar belakang penelitian ini didapatkan beberapa rumusan masalah sebagai berikut.

1. Bagaimana implementasi algoritma *LightGBM* untuk deteksi dini penyakit jantung?
2. Berapa akurasi, *precision*, *recall*, *f1-score*, dan *AUC-ROC* pada algoritma *LightGBM* setelah dilakukan *hyperparameter tuning* untuk deteksi dini penyakit jantung?

1.3 Batasan Permasalahan

Penelitian yang dilakukan memiliki batasan masalah sebagai berikut.

1. Penggunaan dataset “Heart Disease Dataset” berupa data mentah yang berasal dari *UCI Machine Learning Repository* dan diakses pada tahun 2024.
2. Dataset yang digunakan memiliki 303 data dan 13 *features*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini sebagai berikut.

1. Mengimplementasikan algoritma *LightGBM* untuk deteksi dini penyakit jantung.
2. Mengukur akurasi, *precision*, *recall*, *f1-score*, dan *AUC-ROC* pada algoritma *LightGBM* setelah dilakukan *hyperparameter tuning* untuk deteksi dini penyakit jantung.

1.5 Manfaat Penelitian

Manfaat atas penelitian yang dilakukan adalah jika model diterapkan pada aplikasi, membantu pihak medis membuat keputusan deteksi dini penyakit jantung sehingga pasien mendapatkan penanganan yang tepat.

1.6 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Pendahuluan terdiri dari latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, dan manfaat penelitian. Pada pendahuluan menunjukkan gambaran terhadap ide pokok dan alasan dilakukannya penelitian ini.
- Bab 2 LANDASAN TEORI
Landasan Teori berisi teori-teori yang melandasi penelitian ini seperti penyakit jantung, *machine learning*, *supervised learning*, *ensemble learning*, *boosting*, *LightGBM* dan *Confusion Matrix*.

- Bab 3 METODOLOGI PENELITIAN

Metodologi Penelitian menjelaskan mengenai langkah - langkah yang dilakukan dalam penelitian ini. Bagian ini terdapat flowchart alur penelitian dan algoritma yang digunakan pada penelitian ini

- Bab 4 HASIL DAN DISKUSI

Hasil dan Diskusi menjelaskan hasil penelitian dan analisis dari penelitian yang telah dilakukan. Hasil penelitian dan analisis ditunjukkan dalam bentuk gambar, tabel, dan tulisan yang menjelaskan dengan lengkap hasil dari penelitian ini.

- Bab 5 KESIMPULAN DAN SARAN

Bagian ini merupakan bagian simpulan dan saran dari penelitian yang telah dilakukan.

