

PERBANDINGAN ALGORITMA KLASIFIKASI PADA PASIEN PENYAKIT JANTUNG

COMPARISON OF CLASSIFICATION ALGORITHMS IN HEART DISEASE PATIENTS

Jefrin Waruwu, Abdi Dharma

Teknik Informatika, Universitas Prima Indonesia Medan

Jefrinwaruwu1@gmail.com, abdidharma@unprimdn.ac.id

ABSTRACT

The heart is an essential organ in the human body and functions as a central regulator in the circulatory system, where blood vessels act as its distribution pathways. Despite its vital role, the heart is also an organ that is very susceptible to various diseases. Heart disease is one of the most common chronic diseases worldwide, and early diagnosis is essential for effective management. It is important to detect heart disease early, especially if there are congenital factors, to prevent the risk of serious complications and avoidable death. The study aims to compare five classification algorithms in classifying the presence of heart disease based on patient clinical data. The dataset includes age, gender, type of pain, blood pressure, cholesterol, blood sugar, ECG results, heart rate, angina, old peak, ST segment slope, and class. The dataset consists of 1190 rows of data and the data is then divided into 80% training data and 20% validation data. The algorithms compared are CatBoostClassifier, HistGradientBoostingClassifier, RandomForestClassifier, XGBClassifier, and VotingClassifier. Based on the results of the experiments conducted, the HistGradientBoostingClassifier and VotingClassifier algorithms have optimal accuracy compared to the RandomForestClassifier, XGBClassifier, and CatBoostClassifier algorithms. HistGradientBoostingClassifier and VotingClassifier have an accuracy of 97% and 97% while RandomForestClassifier achieves an accuracy of 96% then XGBClassifier achieves an accuracy of 96%, and finally CatBoostClassifier gets an accuracy of 95%. The HistGradientBoostingClassifier and VotingClassifier algorithms are recommended for use. They are more optimal in classifying heart disease patients and normal patients because they get high accuracy compared to other algorithms. Research emphasizes the importance of choosing the right algorithm for early diagnosis and effective management of heart disease to help reduce the risk of complications and improve the quality of life of patients.

Keywords: Heart, Classification, Heart Disease, Algorithm, Data Science

ABSTRAK

Jantung adalah organ yang esensial dalam tubuh manusia dan berfungsi sebagai pusat pengatur dalam sistem sirkulasi, dimana pembuluh darah bertindak sebagai jalur distribusinya. Meskipun memegang peran vital, jantung juga merupakan organ yang sangat rentan terhadap berbagai penyakit. Penyakit jantung merupakan salah satu penyakit kronis yang umum di seluruh dunia, dan diagnosis dini sangat penting untuk pengelolaan yang efektif. Penting untuk mendeteksi penyakit jantung secara awal, terutama jika ada faktor bawaan, untuk mencegah risiko komplikasi serius dan kematian yang dapat dihindari. Penelitian yang dilakukan bertujuan untuk membandingkan lima algoritma klasifikasi dalam mengklasifikasi keberadaan penyakit jantung berdasarkan data klinis pasien. Dataset yang digunakan mencakup informasi umur, jenis kelamin, jenis nyeri, tekanan darah, kolestrol, gula darah, hasil EKG, detak jantung, angina, oldpeak, kemiringan segmen ST, kelas. Dataset terdiri dari 1190 baris data lalu data tersebut dibagi menjadi 80% data latih dan 20% data validasi. Algoritma yang dibandingkan ialah CatBoostClassifier, HistGradientBoostingClassifier, RandomForestClassifier, XGBClassifier, VotingClassifier. Berdasarkan hasil eksperimen yang dilakukan bahwasannya algoritma HistGradientBoostingClassifier dan VotingClassifier memiliki akurasi yang optimal dibandingkan dengan algoritma RandomForestClassifier, XGBClassifier dan CatBoostClassifier. HistGradientBoostingClassifier dan VotingClassifier memiliki akurasi 97% dan 97% sedangkan RandomForestClassifier mencapai akurasi 96% lalu XGBClassifier mencapai akurasi 96% terakhir CatBoostClassifier mendapatkan akurasi 95%. Algoritma HistGradientBoostingClassifier dan VotingClassifier disarankan untuk digunakan karena lebih optimal dalam mengklasifikasi pasien penyakit jantung dan normal karena mendapatkan akurasi yang tinggi dibandingkan algoritma lainnya. Penelitian menegaskan pentingnya pemilihan algoritma yang tepat untuk diagnosis dini dan pengelolaan penyakit jantung yang efektif, sehingga dapat membantu dalam mengurangi risiko komplikasi dan meningkatkan kualitas hidup pasien.

Kata Kunci: Jantung, Klasifikasi, Penyakit Jantung, Algoritma, Ilmu Data.

PENDAHULUAN

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 Indoensia. 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”.

Rumusan Masalah

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

Tujuan dan Manfaat

Tujuan

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

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 refrensi
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.

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 implmentasikan adalah RandomForestClassifier, CatBoostClassifier, HistGradientBosstingClasssifier, XGBClassifier, VottingClassfier
4. Bahasa pemrograman yang diterapkan adalah Python 3
5. Parameter yang diterapkan sebagai ukuran adalah Normal dan Penyakit Jantung

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*, *Logictic 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

METODE

Jenis Penelitian

Studi ini merupakan jenis penelitian eksperimental. Eksperimen yang dilaksanakan bertujuan untuk melakukan perbandingan lima model klasifikasi untuk pasien penyakit jantung yang mendapatkan akurasi terbaik dalam mengkategorikan data data yang tergolong dalam penyakit jantung dan normal.

Waktu dan Tempat Pelaksanaan

Penelitian yang dilakukan berlangsung selama rentang waktu delapan bulan, dimulai pada Oktober 2023 dan berakhir pada Mei 2024. Pengujian metode yang diusulkan dilakukan di Laboratorium Universitas Prima Indonesia. Dataset yang diimplementasikan kedalam penelitian yang dilakukan dalam bentuk data tabular atau CSV dengan kondisi dari kriteria data normal dan data pasien penyakit jantung yang didapatkan dari kaggle. Berikut adalah penjelasan mengenai jadwal penelitian :

Tabel 1. Jadwal Kegiatan

No	Kegiatan	Bulan							
		Okt	Nov	Des	Jan	Feb	Mar	Apr	Mei
1	Diskusi Topik Penelitian	■							
2	Pencarian Jurnal	■							
3	Pengumpulan Data		■						
4	Pembuatan Proposal			■					
5	Diskusi Proposal				■				
6	Pembuatan Aplikasi					■			
7	Pembuatan Proposal						■		
8	Publikasi Jurnal							■	
9	Diseminasi								■

Prosedur Kerja

Adapun beberapa prosedur kerja dari penelitian yang dilakukan untuk mengklasifikasi pasien penyakit jantung dan normal melalui data data tabular atau CSV sebagai berikut:

1. Data diambil dari kaggle lalu ditampilkan dalam bentuk csv setelah itu data dideskripsikan untuk mengetahui tipe dari data data tersebut, lalu data tersebut akan dikelompokkan

dari parameter yang telah digunakan dan akan mendapatkan data target yang berisi numerik 0(normal), 1(Penyakit jantung). Dataset terdiri dari 1190 baris data yang akan digunakan dan akan dibagi menjadi data latih dan data uji. Data latih terdiri dari 952 serta data uji sekitar 238

2. Lalu dilakukan tahap *pre-processing* data, yang dimana pada tahap yang dilakukan mencakup pencarian data yang hilang, distribusi kategorikal variabel, distribusi numerik variabel, melakukan *outliers* pada data, melakukan *correlation between variables*, *correlation between features*, normalisasi data, dan membagi data menjadi dua yaitu data latih dan data uji sehingga data dapat siap digunakan dalam tahap pengujian
3. Lalu dilakukan tahap pengujian, yang dimana data yang telah siap akan digunakan untuk pengujian yang mencakup *accuracy scores of models with cross-validation*, *comparison of models cross validation scores* dan *hyperparameter optimization*, *accuracy scores* pada model *CatBosst*, *HistGradient*, *RandoForest*, *XGBM*, *VotingClassifier*.
4. Lalu dilakukn tahap evaluasi model yang akan memvisualisasikan hasil *confusion matirx* dan *classification report* untuk setiap model

Alat dan Bahan

Pada penelitian yang dilakukan, program yang dibuat menggunakan pyhton di platform Kaggle Colab. Lalu data yang di implementasikan yaitu data tabular atau csv yang didapatkan dari sumber terbuka yaitu Kaggle dan bisa diakses pada tautan berikut

<https://www.kaggle.com/datasets/mexwell/heart-disease-dataset>.

Dataset berisi data numeric dan kateogrikal dalam bentuk tabular atau csv terdiri dari 1190 data yang akan dibagi menjadi data latih dan data uji, yaitu 952 data latih dan 238 data uji. Untuk kejelasan

lebih lanjut mengenai deskripsi data pada gambar dibawah ini

	age	sex	chest pain type	resting bps	cholesterol	fasting blood sugar	resting ecg	max heart rate	exercise angina	oldpeak	ST slope	target
0	40	1	2	140	289	0	0	172	0	0.0	1	0
1	49	0	3	160	180	0	0	156	0	1.0	2	1
2	37	1	2	130	203	0	1	98	0	0.0	1	0
3	48	0	4	138	214	0	0	108	1	1.5	2	1
4	54	1	3	150	195	0	0	122	0	0.0	1	0

Gambar 1. Dataset Penyakit Jantung dan Normal

HASIL DAN PEMBAHASAN

Hasil

Pada penelitian yang dilakukan tersebut, ada beberapa proses yang harus dilakukan dalam melakukan perbandingan algoritma klasifikasi pada penyakit jantung diantaranya seperti mempersiapkan data, memproses data, melakukan pengujian pada algoritma klasifikasi yang akan digunakan.

Persiapan Data

Pada tahap penelitian yang akan dilakukan adalah mempersiapkan data untuk digunakan pada algoritma klasifikasi. Pertama data tersebut diambil dari kaggle melalui google colab dengan menggunakan API (*Application Programing Interface*). Setelah itu akan di visualisasikan dalam bentuk csv untuk menampilkan berapa banyak data yang ada serta tipe datanya lalu apakah ada data yang duplikat dan data yang kosong. Dataset tersebut berupa csv serta tersedia secara terbuka sebagai bahan eksperimen dalam penelitian yang akan dilakukan serta bisa diakses melalui link <https://www.kaggle.com/datasets/mexwell/heart-disease-dataset>. Data yang akan dijadikan eksperimen terdiri dari data numerik dan data kategorikal yang berjumlah 1190 baris data, yang dimana data tersebut akan dibagi menjadi data *validation* dan *training* untuk nantinya akan diujikan pada algoritma klasifikasi.

Pra-pemrosesan Data

Setelah selesai mempersiapkan data, maka selanjutnya adalah memproses data

yang telah dipersiapkan agar algoritma klasifikasi dapat melakukan pengujian dengan baik pada data. Salah satu langkah penting ialah memahami distribusi variabel kategorikal. Distribusi variabel kategorikal membantu memahami bagaimana data tersebar dalam kategori yang berbeda, sehingga kita bisa mengetahui apakah ada ketidakseimbangan atau pola tertentu yang perlu diperhatikan. Penting untuk mengetahui berapa banyak data yang termasuk dalam setiap kategori seperti halnya variabel data kategorikal ialah **sex, chest pain type, fasting blood sugar, resting ecg, exercise angina, ST slope, target**.

Setelah melakukan distribusi variabel kategorikal, langkah selanjutnya melakukan analisis distribusi variabel numerik. Distribusi variabel numerik membantu memahami bagaimana nilai-nilai numerik tersebar dalam dataset, sehingga bisa mengetahui pola, outlier, dan karakteristik lainnya. Variabel numerik pada penelitian ialah **age, resting bp s, cholesterol, max heart rate, oldpeak**.

Selanjutnya menghapus data *outlier* pada data, dengan menghapus *outlier* peneliti memastikan bahwa data yang digunakan untuk melatih model lebih bersih dan representatif. Sehingga membantu meningkatkan akurasi dan keandalan model klasifikasi. Setelah menghapus *outlier* pada dataset, langkah selanjutnya adalah menganalisa korelasi antara variabel. Korelasi membantu memahami hubungan antara dua variabel numerik, apakah mereka memiliki hubungan positif, negatif, atau tidak ada hubungan sama sekali.

Pengujian

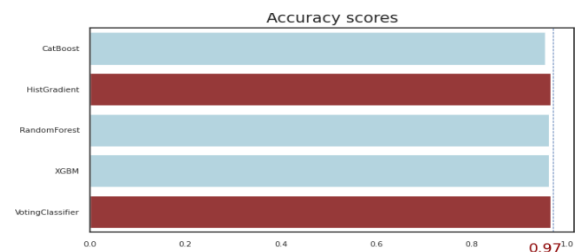
Setelah menyelesaikan pra-pemrosesan data termasuk juga memahami data distribusi variabel kategorikal dan numerik, menghapus *outlier*, serta menganalisis korelasi antara variabel. Setelah tahap pra-pemrosesan data sudah selesai, selanjutnya melakukan *encoding variabel* kategorikal menggunakan *one-hot*

encoding. *One-hot encoding* adalah metode untuk mengubah variabel kategorikal menjadi satu atau lebih variabel biner (*dummy variables*). Dalam *one-hot encoding*, setiap kategori unik dari variabel kategorikal menjadi kolom baru dengan nilai 1 atau 0. Setelah melakukan *one-hot encoding* pada data selanjutnya membagi data menjadi 80% data latih dan 20% data validasi. Dimana data yang sudah dibagi tersebut akan dilakukan normalisasi data agar meningkatkan kinerja dari algoritma model tersebut sehingga algoritma model dapat melakukan pengujian data menjadi lebih baik. Dalam hal ini pengujian menggunakan Optuna untuk mengoptimalkan hiperparameter model dalam melakukan tugas klasifikasi. Berikut visualisasi dari hasil akurasi dari pengujian menggunakan Optuna.

```
CatBoostClassifier 0.9533080395004746
HistGradientBoostingClassifier 0.961010949151542
RandomForestClassifier 0.961064406779061
XGBClassifier 0.961064406779061
VotingClassifier 0.961010949151542
```

Gambar 2. Hasil Pengujian Model Klasifikasi

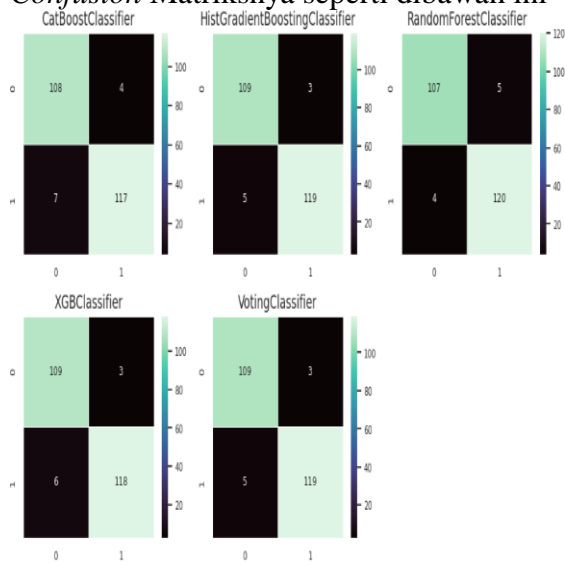
Lalu dibawah ini adalah hasil dari perbandingan visualisasi akurasi pada model agar dapat ditentukan model yang mana yang memberikan akurasi yang lebih bagus dalam melakukan pengujian pada data yang telah digunakan.



Gambar 3. Perbandingan Nilai Akurasi Model

Bisa dilihat pada gambar perbandingan nilai akurasi tersebut bahwasannya model algoritma HistGradient dan VotingClassifier mendapatkan akurasi yang tinggi yaitu sekitar 97%. Selanjutnya akan divisualisasikan jumlah data yang dapat diklasifikasikan berdasarkan data target

yaitu kategori pasien penyakit jantung dan pasien yang normal. Pasien normal dibuat menjadi angka 1 dan pasien penyakit jantung dibuat angka 0 Visualisasi dari Confusion Matriksnya seperti dibawah ini



Gambar 4. Confusion Matrix Algoritma Klasifikasi

Pada gambar menampilkan lima matriks kebingungan yang membandingkan kinerja lima model klasifikasi berbeda: CatBoostClassifier, Hist GradientBoostingClassifier, RandomForestClassifier, XGBClassifier, dan VotingClassifier. Matriks kebingungan ini digunakan untuk mengevaluasi kinerja model klasifikasi dengan menampilkan jumlah prediksi yang benar dan salah untuk masing-masing kelas. CatBoostClassifier menunjukkan hasil dengan 108 prediksi benar untuk kelas negatif, 4 prediksi salah untuk kelas negatif, 117 prediksi benar untuk kelas positif, dan 7 prediksi salah untuk kelas positif. HistGradientBoostingClassifier menghasilkan 109 prediksi benar untuk kelas negatif, 3 prediksi salah untuk kelas negatif, 119 prediksi benar untuk kelas positif, dan 5 prediksi salah untuk kelas positif. RandomForestClassifier menunjukkan kinerja dengan 107 prediksi benar untuk kelas negatif, 5 prediksi salah untuk kelas negatif, 120 prediksi benar untuk kelas positif, dan 4 prediksi salah untuk kelas positif. XGBClassifier menghasilkan 109 prediksi benar untuk

kelas negatif, 3 prediksi salah untuk kelas negatif, 118 prediksi benar untuk kelas positif, dan 6 prediksi salah untuk kelas positif.

Terakhir, VotingClassifier menunjukkan hasil dengan 109 prediksi benar untuk kelas negatif, 3 prediksi salah untuk kelas negatif, 119 prediksi benar untuk kelas positif, dan 5 prediksi salah untuk kelas positif. Secara umum, semua model menunjukkan kinerja yang baik dengan mayoritas prediksi yang benar baik untuk kelas positif maupun negatif. Di antara kelima model tersebut, RandomForestClassifier tampak memberikan hasil terbaik dengan jumlah prediksi benar tertinggi untuk kelas positif (120) dan jumlah prediksi salah terendah untuk kelas positif (4). 0 itu adalah pasien normal dan 1 adalah pasien yang terkena penyakit jantung.

Pembahasan

Setelah berhasil melakukan pengujian model dan memvisualisasikan confusion matrix pada model klasifikasi. Selanjutnya akan mengevaluasi hasil laporan perbandingan akurasi tersebut dibawah ini

Tabel 2. Perbandingan Akurasi Model

Model	Label	Accuracy	Precision	Recall	F1-Score	Support
CatBoostClassifier	0	0.95	0.94	0.96	0.95	112
	1		0.97	0.94	0.96	124
HistGradientBoosting Classifier	0	0.97	0.96	0.97	0.96	112
	1		0.98	0.96	0.97	124
RandomForestClassifier	0	0.96	0.96	0.96	0.96	112
	1		0.97	0.94	0.96	123
XGBClassifier	0	0.96	0.95	0.97	0.96	112
	1		0.98	0.95	0.96	124
VotingClassifier	0	0.97	0.96	0.97	0.96	112
	1		0.98	0.96	0.97	124

Tabel menampilkan tabel perbandingan akurasi untuk lima model klasifikasi berbeda CatBoostClassifier, HistGradientBoosting Classifier, RandomForestClassifier, XGBClassifier

dan VotingClassifier. Perbandingan yang ditampilkan meliputi precision, recall, f1-score, dan support untuk masing-masing kelas (0 dan 1), serta akurasi, rata-rata makro, dan rata-rata tertimbang untuk keseluruhan data. CatBoostClassifier menunjukkan precision, recall, dan f1-score yang seimbang di angka 0.94 hingga 0.96

untuk kedua kelas, dengan akurasi keseluruhan 0.95. HistGradientBoostingClassifier memiliki precision, recall, dan f1-score yang sedikit lebih tinggi, berkisar antara 0.96 hingga 0.97, dan akurasi keseluruhan 0.97. RandomForestClassifier juga menunjukkan metrik yang kuat, dengan precision, recall, dan f1-score antara 0.96 hingga 0.97, serta akurasi 0.96. XGBClassifier memiliki precision, recall, dan f1-score yang hampir serupa dengan model lainnya, dengan angka antara 0.96 hingga 0.98, dan akurasi 0.97. Terakhir, VotingClassifier menunjukkan hasil yang sangat baik dengan precision, recall, dan f1-score antara 0.96 hingga 0.98, serta akurasi keseluruhan 0.97. Secara keseluruhan, semua model menunjukkan kinerja yang baik dengan metrik yang sangat mendekati satu sama lain, dengan HistGradientBoostingClassifier dan VotingClassifier menonjol karena nilai akurasi dan f1-score yang tinggi, menunjukkan kemampuan mereka dalam memberikan prediksi yang sangat akurat dan konsisten.

SIMPULAN

Pada penelitian yang dilakukan mengklasifikasi data pasien penyakit jantung dengan menggunakan lima algoritma klasifikasi yaitu CatBoostClassifier, HistGradientBoostingClassifier, RandomForestClassifier, XGBClassifier, dan VotingClassifier. Dari pemaparan dan pengujian yang dilakukan, dapat dilihat bahwasannya algoritma HistGradientBoostingClassifier dan VotingClassifier memiliki akurasi yang optimal dengan mencapai akurasi 97% dan 97% dari pada algoritma lain yang juga ikut diuji CatBoostClassifier dengan akurasi 95%, RandomForestClassifier memiliki akurasi 96%, dan XGBClassifier juga 96%. Maka dari itu akurasi dari model HistGradientBoostingClassifier dan VotingClassifier lebih optimal

dibandingkan dengan model lain yang telah di uji pada data.

Saran

Setelah menyelesaikan penelitian tersebut, Beberapa saran yang dapat dilakukan untuk pengembangan selanjutnya sebagai berikut:

1. Lakukan *tuning hyperparameter* yang lebih mendalam untuk masing-masing model menggunakan teknik seperti *Grid Search* atau *Random Search*.
2. Melakukan Eksplorasi dan ciptakan fitur-fitur baru yang mungkin lebih informative. Teknik-teknik seperti PCA (*Principal Component Analysis*) atau LDA (*Linear Discriminant Analysis*) dapat digunakan untuk mengurangi dimensi dan menangani korelasi antar fitur.
3. Menggunakan model lain selain model yang telah digunakan pada penelitian yang dilakukan. Model lain seperti halnya *Neural Networks*, SVM(*Support Vector Machine*), atau *Logistic Regression* untuk melihat bagaimana model tersebut berkinerja pada dataset tersebut.

DAFTAR PUSTAKA

- [1] dr. Sienny Agustin, "Memahami Sistem Peredaran Darah Manusia Beserta Fungsinya," *Alodokter*, 2022.
<https://www.alodokter.com/memahami-sistem-peredaran-darah-pada-manusia>
- [2] Bunda, "Kenali Fungsi Kerja Jantung Anda," *RSU Bunda*, 2021.
<https://bunda.co.id/artikel/kesehatan/jantung/kenali-fungsi-kerja-jantung-anda/>
- [3] Rokom, "Penyakit Jantung Penyebab Utama Kematian, Kemenkes Perkuat Layanan Primer," *Redaksi Sehat Negeriku*, 2022.
<https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220929/0541166/penyakit->

- jantung-penyebab-utama-kematian-kemenkes-perkuat-layanan-primer/
- [4] A. Riski, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [5] A. Purnama, "Edukasi Dapat Meningkatkan Kualitas Hidup Pasien yang Terdiagnosa Penyakit Jantung Koroner," *J. Kesehat. Indones. Indones. J. Heal.*, vol. X, no. 2, pp. 66–71, 2020.
- [6] A. Ahdiat, "Kasus Penyakit Katastropik di Indonesia Meningkat pada 2022," *databoks*, 2023. <https://databoks.katadata.co.id/datapublish/2023/07/03/kasus-penyakit-katastropik-di-indonesia-meningkat-pada-2022> (accessed Aug. 26, 2023).
- [7] Eva Miranda Marwali, Yoel Purnama, and Poppy Surwianti Roebiono, "Modalitas Deteksi Dini Penyakit Jantung Bawaan di Pelayanan Kesehatan Primer," *J. Indones. Med. Assoc.*, vol. 71, no. 2, pp. 100–109, 2021, doi: 10.47830/jinma-vol.71.2-2021-241.
- [8] A. B. Wibisono and A. Fahrurrozi, "Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner," *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2019, doi: 10.35760/tr2019.v24i3.2393.
- [9] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, "Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung," *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 189–194, 2019, doi: 10.33480/pilar.v15i2.6
- [10] F. Handayani *et al.*, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. Vol. 7 No. 3, 2021.
- [11] M. Balakrishnan, A. B. Arockia Christopher, P. Ramprakash, and A. Logeswari, "Prediction of Cardiovascular Disease using Machine Learning," *J. Phys. Conf. Ser.*, vol. 1767, no. 1, pp. 1–9, 2021, doi: 10.1088/1742-6596/1767/1/012013.
- [12] J. Junifer Pangaribuan, H. Tanjung, and K. Kenichi, "Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression," *J. Inf. Syst. Dev.*, vol. 06, no. 02, pp. 1–10, 2021.
- [13] H. Hidayat, A. Sunyoto, and H. Al Fatta, "Klasifikasi Penyakit Jantung Menggunakan Random Forest Clasifier," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 31–40, 2023, doi: 10.47970/siskom-kb.v7i1.464.
- [14] A. Sefarni, I. E. Hendrawan, and C. Rozikin, "Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4.5," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 7, no. 2, p. 117, 2022, doi: 10.30998/string.v7i2.12012.
- [15] D. Nasien, R. Syahputra, A. Akbar Marunduri, and R. Prawinata See, "Klasifikasi Penyakit Jantung Menggunakan Decision Tree dan KNN Menggunakan Ekstraksi Fitur PCA," vol. 4, no. 1, pp. 1–6, 2024.
- [16] J. D. Muthohhar and A. Prihanto, "Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung," *J. Informatics Comput. Sci.*, vol. 04, pp. 298–304, 2023, doi: 10.26740/jinacs.v4n03.p298-304.
- [17] A. S. Prabowo and F. I. Kurniadi, "Analisis Perbandingan Kinerja

- Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 7, no. 1, pp. 56–61, 2023, doi: 10.47970/siskom-kb.v7i1.468.
- [18] I. M. Krisna, D. Jaya, I. G. Agung, and G. Arya, “Perbandingan Random Forest , Decision Tree , Gradient Boosting , Logistic Regression untuk Klasifikasi Penyakit Jantung,” vol. 2, no. November, pp. 61–70, 2023.
- [19] A. Masruriyah, H. Novita, C. Sukmawati, A. Ramadhan, S. Arif, and B. Dermawan, “Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung,” *Comput. Sci.*, vol. 4, no. 1, pp. 62–70, 2024, doi: 10.31294/coscience.v4i1.2389.
- [20] Ratnasari, A. Jurnaidi Wahidin, A. Eko Setiawan, and P. Bintoro, “Machine Learning Untuk Klasifikasi Penyakit Jantung,” *Aisyah J. Informatics Electr. Eng.*, vol. 6, no. 1, pp. 145–150, 2024, doi: 10.30604/jti.v6i1.272.
- [21] N. Yudistira and A. F. Putra, “Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance,” *J. Litbang Edusaintech*, vol. 2, no. 2, pp. 112–122, 2021, doi: 10.51402/jle.v2i2.48.