



Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree

Mochammad Ilham Aziz*, Ahmad Zainul Fanani, Affandy

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia

Email: ¹*il.ilham95@gmail.com, ²a.zainul.fanani@dsn.dinus.ac.id, ³affandy@dsn.dinus.ac.id

Email Penulis Korespondensi : il.ilham95@gmail.com

Abstrak—Penyakit jantung ialah salah satu penyakit yang tidak mudah di prediksi semenjak dini dan mungkin Sebagian orang tidak sadar bahwa mengidap penyakit tersebut. Data yang diperoleh WHO Lebih dari 17 juta orang di seluruh dunia meninggal karena serangan jantung pada tahun 2016. Jika gejala penyakit jantung atau serangan jantung diketahui maka dapat dilakukan antisipasi pencegahan terhadap penyakit jantung bahkan dapat meminimalkan angka kematian. Analisis terhadap penyakit jantung bertujuan untuk mengurangi angka kematian terhadap penyakit tersebut. Pada penulisan penelitian ini digunakan suatu metode algoritma decision tree, Algoritma tersebut masih memiliki kelemahan dalam melakukan keakuratan prediksi. Maka membutuhkan suatu cara untuk meningkatkan akurasi dari hasil pembelajaran klasifikasi. Penelitian ini bertujuan untuk meningkatkan hasil pembelajaran klasifikasi dari penyakit jantung dengan menggunakan metode pembelajaran ensemble yaitu Bootstrap Aggregating (Bagging) dan Adaptive Boosting (Adaboost). Kedua metode tersebut diuji dengan melakukan prediksi kematian yang disebabkan oleh penyakit jantung.

Kata Kunci: Decision Tree; Penyakit Jantung; Kematian; Bagging; Adaboost

Abstract—Heart disease is one disease that is not easy to predict early on and maybe some people are not aware that they have the disease. Data obtained by WHO More than 17 million people worldwide died of heart attacks in 2016. If the symptoms of heart disease or heart attack are known, prevention of heart disease can be anticipated and even minimized mortality. Analysis of heart disease aims to reduce mortality from the disease. In writing this research, a decision tree algorithm method is used, the algorithm still has weaknesses in making prediction accuracy. So we need a way to improve the accuracy of the classification learning outcomes. This study aims to improve the learning outcomes of heart disease classification by using ensemble learning methods, namely Bootstrap Aggregating (Bagging) and Adaptive Boosting (Adaboost). Both methods were tested by predicting deaths caused by heart disease.

Keywords: Decision Tree; Heart Disease; Death; Bagging; Adaboost

1. PENDAHULUAN

Tahapan-tahapan ini melakukan penyusunan dan juga paparan daripada pengetahuan yang telah sudah didapat dari data mining sebelumnya. Representasi daripada hasil data mining tersebut, dibuatlah sedemikian bentuknya agar orang yang masih pemula atau awam dalam dunia data mining dapat mengerti dan juga membaca representasinya. antisipasi siapapun, termasuk anak-anak, bisa terkena penyakit jantung. Dengan memprediksi risiko penyakit jantung dapat mengetahui seseorang terkena penyakit jantung atau tidak dari gejala yang ditimbulkan.

Tekanan darah tinggi, kolesterol tinggi, stres, ketegangan, konsumsi alkohol, gaya hidup sedentari, obesitas, diabetes adalah faktor utama yang mempengaruhi jantung. Tekanan darah tinggi dan kolesterol faktor yang mempengaruhi seseorang mengidap penyakit jantung [3].

Data mining merupakan cara untuk menemukan relasi yang bermakna, bentuk serta tren yang cenderung memeriksa jumlah data yang disimpan memakai teknik pola pengenalan dan dalam data mining ada metode biasa disebut klasifikasi. Ada banyak algoritma klasifikasi, Satu diantara metode klasifikasi yang paling populer dan mudah untuk dipahami dan sering digunakan adalah decision tree [10,14,15].

Pohon keputusan merupakan bentuk pohon seperti diagram alur, dimana tiap titik internalnya mewakili uji karakter. Hasil uji terwakili oleh cabang selain itu label kelas juga terwakili oleh simpul daun. Penilaian uji terhadap variable atribut X pada pohon keputusan Jalur pada akar atau simpul daun memegang prediksi kelas dari tupel yang dilacak. Model pembelajaran prediktif digunakan untuk pengambilan keputusan yang mampu memetakan observasi terhadap suatu kesimpulan berdasarkan target nilai. Nilai ini merupakan metode model prediktif yang sering dipakai di statistic, data mining, dan machine learning.

Pada bentuk pohon, label kelas terwakili oleh daun serta label kelas terwakili oleh cabang yang terkoneksi fitur [7]. Dalam pendekatan pohon keputusan, pruning adalah proses eliminasi Sebagian cabang yang tidak diperlukan. Metode pemangkasan selanjutnya dengan menggunakan ambang batas (threshold pruning). Kedua, pengujian dengan memakai metode pruning mampu mengklasifikasi pohon keputusan dengan syarat segmentasi dari algoritma C4.5 dan ID3 (biaya heterogen) pada enam dataset dan menyimpulkan bahwa, keduanya bisa dipakai sebagai pengklasifikasi pohon keputusan.

Ada jenis model klasifikasi terkait dengan memprediksi stroke antara lain Tsukamoto, SVM, K-NN, Naïve Bayes (NB), Neural Network and Decision Tree (C4.5) [13],[14]. Dari berbagai penelitian tersebut model Decision Tree adalah yang tertinggi untuk tingkatan akurasinya yaitu sebesar 95%. Oleh karenanya metode Decision Tree juga terdapat keunggulan seperti mampu memberikan solusi berbagai jenis overfitting, penanganan missing value atribut di data latih dan juga bisa meningkatkannya efisiensi pada komputasi. Walaupun model C4.5 atau Decision



Tree lebih unggul dari model yang lain disamping itu model ini pada kasus penyakit stroke tidak bisa memprediksi secara tepat dan akurat [16].

Oleh karenanya dalam mengatasi masalah yang telah ada Sebagian peneliti sudah menerapkan teknik yang disebut ensemble yang dianggap mampu dapat memperbaiki nilai prediksi yang tidak tepat dapat juga meningkatkannya hasil akurasi dari Model Decision Tree yang akurat untuk prediksi didalam penelitian [17]. Berdasarkan uraian permasalahan diatas, metode ensemble adalah tujuan utamanya penelitian ini dikarenakan dapat meningkatkan hasil akurasi. Maka klasifikasi penyakit jantung masih bisa ditingkatkan kinerjanya dengan menggunakan Algoritma C4.5 dengan teknik ensemble dipilih guna membantu memprediksi screening pasien penyakit jantung.

Penelitian ini bertujuan untuk melakukan perbandingan antara algoritma Bootstrap Aggregating (Bagging) dan Adaptive Boosting (AdaBoost) dalam meningkatkan hasil akurasi klasifikasi pada data penyakit jantung dan untuk mengevaluasi metode ensemble yang tepat untuk klasifikasi penyakit jantung.

2. METODOLOGI PENELITIAN

2.1 Data Eksperimen

Objek yang diteliti adalah tahapan dalam penyakit jantung. Data yang dikumpulkan berasal dari dataset deteksi penyakit jantung dari website The Kaggle Learning Repository. Kumpulan data ini berasal dari tahun 1988 dan terdiri dari empat database: Cleveland, Hungaria, Swiss, dan Long Beach V. Dataset ini diambil dari website The Kaggle Learning Repository karena website tersebut sudah sering digunakan sebagai sumber dataset oleh pelajar, edukator dan peneliti lainnya.

Dataset penyakit jantung ini berisikan indikator-indikator atau variabel yang dipakai dalam mengidentifikasi pasien. Dataset ini sudah dipisahkan dan dipilih dari variabel-variabel pada dataset penyakit jantung awal dan dibuat menjadi dataset penyakit jantung baru.

Tabel 1. Heart disease dataset

No	Atribut	Representasi	Keterangan	Tipe Data	Deskripsi
1	Age	Age	Umur	Numeric	Age in years
2	Sex	Sex	Jenis kelamin	Categorical	1 : Male 2 : Female
3	Cp	Chest pain type	Nyeri pada dada	Categorical	1 : Typical Angina 2 : Atypical Angina 3 : non angina pain 4 : asymptomatic
4	Trestbps	resting blood pressure	Tekanan darah pasien	Numeric	Mm/hg
5	Chol	serum cholestoral	Kadar kolestrol	Numeric	Mg/dl
6	Fbs	fasting blood sugar	Kadar gula darah	Categorical	1 : True 2 : False
7	Restecg	Resting electrocardiographic result	Hasil tes elektrokardiografi	Categorical	0 : Normal 1: Abnormal 2 : Probable
8	Thalach	heart rate achieved	Denyut jantung	Numeric	Detak jantung
9	Exang	exercise induced angina	Nyeri saat olahraga	Categorical	1 : Ya 0 : tidak
10	Oldpeak	depression induced by exercise relative to rest	Penurunan ST disebabkan berolahraga	Numeric	Depresi Induksi olahraga
11	Slope	Slope	Slope puncak ST Setelah berolahraga	Categorical	1 : upsloping 2 : flat 3: downsloping
12	Ca	The number of blood vessels detected	Pembulu darah	Numeric	Jumlah pembuluh
13	Thal	Thallium	Thallium	Categorical	1 : Normal 2 : Fixed 3: reversible
14	Class	Class	Terkena penyakit jantung	Categorical	1 : Ya 0 : tidak



Tabel 2. Heart disease prediction

No	Atribut	Representasi	Keterangan	Tipe Data	Deskripsi
1	Age	Age	Umur	Numeric	Angka
2	Sex	Sex	Jenis kelamin	Categorical	1 : Male 2 : Female
3	Cp	Chest pain type	Nyeri pada dada	Categorical	1 : Typical Angina 2 : Atypical Angina 3:non angina pain 4: asytmomatic
4	Bp	resting blood pressure	Tekanan darah pasien	Numeric	Mm/hg
5	Cholestrol	serum cholestoral	Kadar kolestrol	Numeric	Mg/dl
6	Fbs over 120	fasting blood sugar	Kadar gula darah	Categorical	1 : Ya 0 : tidak
7	EKG result	EKG Result Patient	Hasil EKG Pasien	Nominal	0 : ekgmo 1 : ekgdy 2 : ekgyr
8	Max HR	Maximum Heart Rate	Detak jantung	Numeric	Angka
9	Exercise Angina	exercise induced angina	Nyeri saat olahraga	Categorical	1 : Ya 0 : tidak
10	ST Depression	depression induced by exercise relative to rest	Penurunan ST Disebabkan berolahraga	Numeric	Depresi Induksi Olahraga
11	Slope of ST	Slope	Slope puncak ST setelah berolahraga	Categorical	1 : upsloping 2 : flat 3 : downsloping
12	Number of vessel fluoro	The number of blood vessels detected through the staining process	Pembulu darah	Numeric	Jumlah Pembulu Darah
13	Thallium	Thallium	Kondisi Tallium	Categorical	1 : Normal 2 : Fixed 3 : reversable
14	Heart Disease	Heart Disease	Penyakit Jantung	Categorical	1 : Absence 0 : Presence

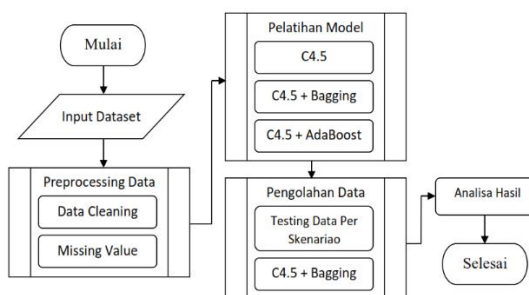
Tabel 3. Heart failure prediction

No	Atribut	Representasi	Keterangan	Tipe Data	Deskripsi
1	Age	Age	Umur	Numeric	Angka
2	Anemia	Decrease of red blood cells or hemoglobin	Pasien anemia	Categorical	1 : Ya 0 : Tidak
3	Ceratinine phosphokinase	Level of the CPK enzyme in the blood	Enzim dalam darah	Numeric	Mcg/L
4	Diabetes	If the patient has diabetes	Diabetes pasien	Categorical	1 : Ya 0 : Tidak
5	Ejection Fraction	Percentage of blood leaving the heart at each contraction	Presentase dalam darah	Numeric	Angka
6	High Blood Pressure	If the patient Has hypertension	Tekanan Darah	Categorical	1 : Ya 0 : Tidak
7	Platelets	Platelets in the blood	Kadar Trombosit	Numeric	Kiloplatelets/ mL
8	Serum	Level of serum	Kadar Kreatin	Numeric	mg/dL



No	Atribut	Representasi	Keterangan	Tipe Data	Deskripsi
9	ceratine Serum Sodium	creatinine in the blood Level of serum sodium in the blood	Kadar Sodium	Numeric	mEq/L
10	Sex	Sex	Jenis kelamin	Numeric	Angka
11	Smoking	If the patient smokes or not	Pasien perokok	Categorical	1 : Ya 0 : Tidak
12	Time	Follow-up period	Waktu tindak lanjut	Numeric	Angka
13	Death Event	If the patient follow-up period	Kematian pasien	Nominal	1 : Ya 0 : Tidak

2.2 Alur Metode



Gambar 1. Alur Metode

2.3 Metode Evaluasi

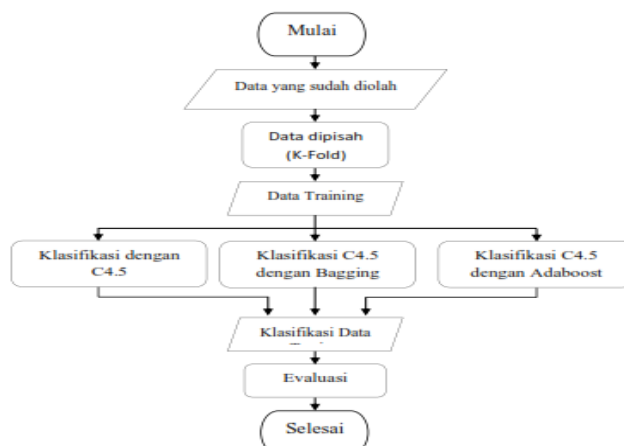
Tahapan selanjutnya yaitu preprocessing setelahnya selesai, dapat melanjutkan pada tahap pelatihan model. Memasuki tahap ketiga akan dilakukan uji klasifikasi antara algoritma Decision Tree (C4.5), C4.5 dan Bagging, C4.5 dan AdaBoost. Tujuan ini dilakukan dengan cara membagi proses klasifikasi menjadikan tiga tahapan ataupun skenario.

Tahapan-tahapan ini melakukan penyusunan dan juga paparan daripada pengetahuan yang telah sudah didapat dari data mining sebelumnya. Representasi daripada hasil data mining tersebut, dibuatlah sedemikian bentuknya agar orang yang masih pemula atau awam dalam dunia data mining dapat mengerti dan juga membaca representasinya.

Tabel 4. Tahapan klasifikasi

Algoritma C4.5	Algoritma C4.5 + Bagging	Algoritma C4.5 + Adaboost
Metode ensemble di tahap belum diterapkan hanya menggunakan algoritma C4.5 saja.	Metode ensemble di tahap ini akan menggunakan metode Bagging untuk mendapatkan hasil klasifikasi yang akurat.	Metode ensemble di tahap ini akan menggunakan metode AdaBoost untuk mendapatkan hasil klasifikasi yang akurat.

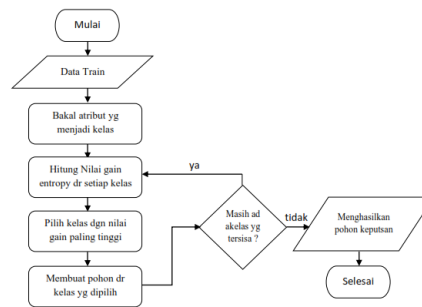
Berdasarkan skenario sebelumnya, langkah -langkah dalam proses klasifikasi dapat diilustrasikan ke dalam flowchart seperti Gambar dibawah ini:



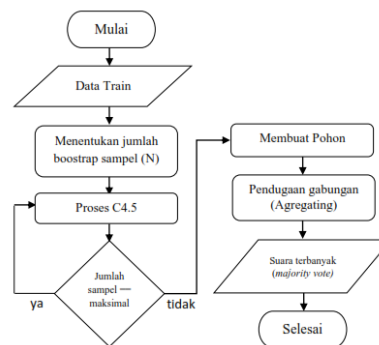
Gambar 2. Flowchart Proses Klasifikasi



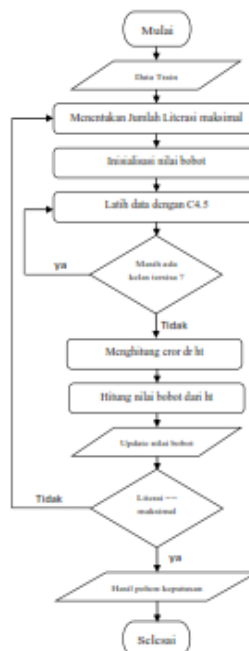
1. Flowchart Klasifikasi Algoritma C4.5

**Gambar 3.** Flowchart Klasifikasi C4.5

2. Flowchart Klasifikasi algoritma C4.5 + Bagging

**Gambar 4.** Flowchart Klasifikasi C4.5 dengan Bagging

3. Flowchart Klasifikasi Algoritma C4.5 + Adaboost

**Gambar 5.** Flowchart klasifikasi C4.5 dengan AdaBoost**2.3 Seting Eksperimen**

pada tahapan seting eksperimen merupakan tahapan akhir dari penelitian adalah model uji dari masing-masing ketiga skenario klasifikasi diatas, yang mana data digunakan sudah mempunyai label tiap datanya. Metode confusion matrix dilakukan untuk model pengujian ini. Sebelum dilakukan pengujian, Pengujian dilakukan yang pertama-tama adalah data harus dipisahkan jadi dua yaitu data training dan data testing. Lalu data bisa diolah dan diuji menggunakan kfold cross-validation dengan cara membagi bagian dengan 10 bagian data secara acak dan stratified untuk mengevaluasi model.



2.4 Analisis Hasil

Pada pengujian sebelumnya mengenai pembahasan komparasi hasil scenario yang terdiri dari algoritma C4.5, algoritma C4.5 + bagging, algoritma C4.5 + adaboost. Selain itu dengan menggunakan metode yang sama, juga dilakukan analisis untuk mengetahui perbandingannya.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan data dari pasien penyakit jantung yang berasal dari website Kaggle (1) Heart Disease Dataset (2) Heart Disease Prediction (3) Heart Failure Prediction.

Tabel 5. Dataset Yang Digunakan Eksperimen

Dataset	Jumlah Record	Jumlah Atribut	Jumlah Atribut Nominal	Jumlah Atribut Numerik	Jumlah Class
Heart Disease Dataset	1025	14	7	7	1
Heart Disease Prediction	270	14	6	8	1
Heart Failure Prediction	300	13	5	8	1

3.2 Pengolahan Data

3.2.1 Data Cleaning

Proses ini proses pembersihan data dapat dilaksanakan terhadap data yang missing value dari ketiga dataset yang digunakan penelitian yaitu Heart Disease Dataset, Heart Disease Prediction dan Heart Failure Prediction.

Tabel 6. Sampel Dataset Heart Disease Setelah Proses Data Cleaning

Age	Sex	Cp	...	Ca	Thal	Class
52	M	Typical	...	0	Defect	1
60	F	Atypical	...	0	Normal	0
:	:	:	:	:	:	:
70	M	Typical	...	0	Normal	0
54	F	Atypical	...	1	Normal	1

Tabel 7. Sampel Dataset Heart Disease Prediction Setelah Proses Data Cleaning

Age	Sex	Bp	...	Slope	Number Vessel	Heart Disease
67	M	115	...	Flat	0	1
65	M	124	...	Upsloping	3	1
:	:	:	:	:	:	:
37	F	120	...	Defect	2	0
45	M	160	...	Flat	1	1

Tabel 8. Sampel Dataset Heart Failure Setelah Proses Data Cleaning

Age	Anemia	Diabetes	...	Sex	Smoking	Death Event
75	0	0	...	M	Yes	Yes
55	0	1	...	M	No	Yes
:	:	:	:	:	:	:
45	1	1	...	F	Yes	No
87	0	1	...	M	No	Yes

3.2.2 Data Transformation

Tabel 9. Sampel Dataset Heart Disease Yang Sudah Dikonversi

Age	Sex	Cp	...	Ca	Thal	Class
52	1	0	...	0	2	1
60	0	1	...	0	1	0
:	:	:	:	:	:	:
70	1	0	...	0	1	0



Age	Sex	Cp	...	Ca	Thal	Class
54	0	1	...	1	1	1

Beberapa atribut yang memang memiliki data-data kategorikal yang sudah diubah menjadi tipe data numerikal. Untuk mempermudahnya melihat secara detail perubahan atribut dari setiap data. Table dibawah ini melihatkan detail apa saja atribut kategorikal yang telah dilakukan perubahan tipe data.

Tabel 10. Detail Atribut Kategorikal Yang Diubah Menjadi Numerik

Atribut Kategorikal	Sebelum dikonversi	Sesudah dikonversi
Sex	Female Male	Female=0 Male=1
Chest Pain Type	Typical Angina Atypical Angina Non-Angina Pain Asymptomatic	Typical Angina=0 Atypical Angina=1 Non-Angina Pain=2 Asymptomatic=3
Fasting Blood Sugar	False True	False=0 True=1
Resting	Normal Having ST-T wave	Normal=0 Having ST-T wave=1
Electrocardiographic	Probable	Probable=2
Exang	Yes No	Yes=1 No=0
Slope	Upsloping Flat Downsloping	Upsloping=0 Flat=1 Downsloping=2
Thal	Normal Fixed Defect Reversable Defect	Normal=1 Fixed Defect=2 Reversable Defect=3
Num	<50% diameter narrowing >50% diameter narrowing	<50%=0 >50%=1

Di dalam dataset Heart Disease Prediction ada sebagian atribut data yang mempunyai tipe data kategorikal antara lain adalah atribut sex, cp (chest pain type), fbs (fasting blood sugar), EKG result, Exercise Angina, slope, thal, Vessel. Table di berikut ini.

Tabel 11. Detail Atribut Heart Disease Prediction Kategorikal Yang Diubah Menjadi Numerik

Atribut Kategorikal	Sebelum dikonversi	Sesudah dikonversi
Sex	Female Male	Female=0 Male=1
Chest Pain Type	Typical Angina Atypical Angina Non-Angina Pain Asymptomatic	Typical Angina=0 Atypical Angina=1 Non-Angina Pain=2 Asymptomatic=3
Fasting Blood Sugar	False True	False=0 True=1
EKG Result	exercise ECG reading	Normal=0 Having ST-T Wave=1 Probable=2
Exercise Angina	Yes No	Yes=1 No=0
Slope	Upsloping Flat Downsloping	Upsloping=0 Flat=1 Downsloping=2
Thal	Normal Fixed Defect Reversable Defect	Normal=1 Fixed Defect=2 Reversable Defect=3
Vessel	Number Of Major Vessel	0-3 flouroscopy

Di dalam dataset Heart Failure ada sebagian atribut data yang mempunyai tipe data kategorikal antara lain adalah atribut sex, Diabetes, Blood Preasure, Smoking, Death Event. Table di berikut ini.



Tabel 12. Detail Atribut Heart Failure Kategorikal Yang Diubah Menjadi Numerik

Atribut Kategorikal	Sebelum dikonversi	Sesudah dikonversi
Sex	Female	Female=0
	Male	Male=1
Diabetes	History of Diabetes	History of Diabetes=1
	No Such History	No Such History=0
Blood Preasure	True	True=1
	False	False=0
Smoking	Yes	Yes=1
	No	No=0
Death Event	Yes	Yes=1
	No	No=0

3.2.3 Klasifikasi Algoritma C4.5

Melihat dari sekenario klasifikasi table diatas, model pelatihan awal adalah mencoba menggunakan metode algortima C4.5. tahapan kalsifikasi akan diolah menggunakan software RapidMiner. Tahapan tersebut ketika proses pengklasifikasian mempunyai tahapan yaitu dengan dua tahapan saat prosesnya dengan data training dan testing. Agar menemukan hasil output terbaik maka maka tahapan tersebut harus dilakukan pada algoritma C.45.

3.2.4 Klasifikasi Algoritma C4.5 + Bagging

Melihat dari sekenario klasifikasi table diatas, model pelatihan kedua adalah mencoba menggunakan metode algortima C4.5 dengan ensemble Bagging. Tahapan tersebut ketika proses pengklasifikasian mempunyai tahapan yaitu dengan dua tahapan saat prosesnya dengan data training dan testing. Agar menemukan hasil output terbaik maka maka tahapan tersebut harus dilakukan pada algoritma C4.5 dengan Bagging.

3.2.5 Klasifikasi Algoritma C4.5 + Adaboost

Proses pengkalsifikasian untuk data latih harus dilakukan dengan cara mengulang hingga memenuhi batas maksimal yang telah ditentukan. Lalu setelahnya hasil dari pengulangan itu akan dipilih dari pengambilan suara yang paling banyak untuk memperoleh keluaran yang terbaik. Software RapidMiner masih menjadi alat untuk mengolah proses klasifikasi kali ini dan sama dengan metode sebelum-sebelumnya.

3.2.6 Uji Validasi dan Evaluasi

1. Algoritma C4.5

Metode confusion matrix diterapkan pada pengujian model ini, untuk mengetahui nilai dari kelas yang sebenarnya dan nilai kelas prediksi dari Dataset Penyakit Jantung.

Tabel 13. Confusion Matrix Dari Algoritma C4.5 Dataset Heart Disease

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	417	74	0.84
	Jantung	82	452	0.84
Class Recall		0.83	0.85	
Class F1Score		0,83	0.84	

Diketahui :

Benar Positif = 417

Benar Negatif = 452

Salah Positif = 74

Salah Negatif = 82

Tabel 14. Confusion Matrix Dari Algoritma C4.5 Dataset Heart Disease Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	81	34	0.70
	Jantung	39	116	0.74
Class Recall		0.67	0.77	
Class F1Score		0.69	0.74	

Diketahui :

Benar Positif = 81

Benar Negatif = 116

Salah Positif = 34



Salah Negatif = 39

Tabel 15. Confusion Matrix Dari Algoritma C4.5 Dataset Heart Failure Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	56	20	0.73
	Jantung	40	183	0.82
Class Recall		0.58	0.90	
Class F1Score		0.65	0.81	

Diketahui :

Benar Positif = 56

Benar Negatif = 183

Salah Positif = 20

Salah Negatif = 40

3.2.7 Algoritma C4.5 + Bagging

Metode confusion matrix diterapkan pada pengujian model ini, untuk mengetahui nilai dari kelas yang sebenarnya dan nilai kelas prediksi dari Dataset Penyakit Jantung.

Tabel 16. Confusion Matrix Dari Algoritma C4.5 + Bagging Dataset Heart Disease

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	440	42	0.91
	Jantung	59	484	0.89
Class Recall		0.88	0.92	
Class F1Score		0.90	0.92	

Diketahui :

Benar Positif = 440

Benar Negatif = 484

Salah Positif = 42

Salah Negatif = 59

Tabel 17. Confusion Matrix Dari Algoritma C4.5 + Bagging Dataset Heart Disease Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	82	30	0.73
	Jantung	38	120	0.75
Class Recall		0.68	0.80	
Class F1Score		0.71	0.76	

Diketahui :

Benar Positif = 82

Benar Negatif = 120

Salah Positif = 30

Salah Negatif = 38

Tabel 18. Confusion Matrix Dari Algoritma C4.5 + Bagging Dataset Heart Failure Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	61	19	0.76
	Jantung	35	184	0.84
Class Recall		0.63	0.90	
Class F1Score		0.69	0.83	

Benar Positif = 61

Benar Negatif = 184

Salah Positif = 19

Salah Negatif = 35

3.2.8 Algoritma C4.5 + Adaboost

Metode confusion matrix juga digunakan pada pengujian kombinasi algoritma C4.5 dan AdaBoost, untuk mengetahui nilai dari kelas yang sebenarnya dan nilai kelas prediksi dari Dataset Penyakit Jantung.



Tabel 19. Confusion Matrix Dari Algoritma C4.5 + Adaboost Dataset Heart Disease

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	433	55	0.88
	Jantung	66	471	0.87
Class Recall		0.86	0.89	
Class F1Score		0.88	0.89	

Diketahui :

Benar Positif = 433

Benar Negatif = 471

Salah Positif = 66

Salah Negatif = 55

Tabel 20. Confusion Matrix Dari Algoritma C4.5 + Adaboost Dataset Heart Disease Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	87	24	0.78
	Jantung	33	126	0.79
Class Recall		0.72	0.84	
Class F1Score		0.75	0.81	

Diketahui :

Benar Positif = 87

Benar Negatif = 126

Salah Positif = 24

Salah Negatif = 33

Tabel 21. Confusion Matrix Dari Algoritma C4.5 + Adaboost Dataset Heart Failure Prediction

Confusion Matrix		Kelas Sebenarnya		Class Presisi
		No Jantung	Jantung	
Kelas Prediksi	No Jantung	56	20	0.73
	Jantung	40	183	0.82
Class Recall		0.58	0.90	
Class F1Score		0.65	0.81	

Diketahui :

Benar Positif = 56

Benar Negatif = 183

Salah Positif = 20

Salah Negatif = 40

3.3 Analisis Hasil

Setelah membahas hasil dari tiap skenario pada ketiga dataset yang digunakan Dataset Heart Disease, Heart Disease Prediction, dan Heart Failure Prediction untuk itu proses setelahnya yaitu melaksanakan analisis hasil dari evaluasi dengan cara membandingkan performa pengklasifikasian. Skenario yang pertama adalah pengklasifikasian dengan metode algoritma C4.5 dimana ketika metode tersebut dieksekusi kedalam software rapidminer memperoleh hasil pada table dibawah ini.

Tabel 22. Hasil Pengujian Skenario Pertama

Uji Validasi	Algoritma C4.5		
	Heart Disease Dataset	Heart Disease Prediction	Heart Failure Prediction
Presisi	0.84	0.70	0.74
Spesifitas	0.85	0.77	0.90
Sensivitas	0.83	0.68	0.58
F1 Score	0.83	0.69	0.65
Akurasi	0.84	0.72	0.79

Skenario selanjutnya merupakan yang kedua adalah pengklasifikasian dengan metode C4.5 dan Bagging. Dimana dalam skenario ini adalah kombinasi dari algoritma C4.5 dan metode Bagging adalah untuk meningkatkan performa klasifikasi dari C4.5 itu sendiri.

**Tabel 23.** Hasil Pengujian Skenario Kedua

Uji Validasi	C4.5 dengan Bagging		
	Heart Disease Dataset	Heart Disease Prediction	Heart Failure Prediction
Presisi	0.91	0.73	0.76
Spesifitas	0.92	0.80	0.91
Sensivitas	0.88	0.68	0.64
F1 Score	0.90	0.71	0.69
Akurasi	0.90	0.74	0.81

Skenario selanjutnya merupakan yang terakhir adalah pengklasifikasian dengan metode C4.5 dan Adaboost. Dimana dalam skenario ini adalah kombinasi dari algoritma C4.5 dan metode Adaboost adalah untuk meningkatkan performa klasifikasi dari C4.5 itu sendiri.

Tabel 24. Hasil Pengujian Skenario Ketiga

Uji Validasi	C4.5 dengan AdaBoost		
	Heart Disease Dataset	Heart Disease Prediction	Heart Failure Prediction
Presisi	0.89	0.78	0.74
Spesifitas	0.90	0.84	0.90
Sensivitas	0.87	0.73	0.58
F1 Score	0.88	0.75	0.65
Akurasi	0.88	0.78	0.79

Berdasarkan Tabel XXIII dan XXIV penerapan metode bagging dan adaboost memiliki pengaruh dalam meningkatnya nilai akurasi dari algoritma C4.5. Akurasi yang dihasilkan oleh kombinasi algoritma dengan metode bagging dan adaboost untuk klasifikasi ketiga Dataset Penyakit Jantung dapat dikatakan lebih baik dibandingkan hanya menerapkan satu metode saja. Tabel XXV menunjukkan detail perbandingan hasil pengujian untuk keseluruhan skenario klasifikasi.

Tabel 25. Perbandingan Hasil Pengujian

Dataset	Uji Validasi Akurasi		
	C4.5	C4.5 + Bagging	C4.5 + AdaBoost
Heart Disease Dataset	84.78%	90.14%	88.20%
Heart Disease Prediction	72.96%	74.81%	78.89%
Heart Failure Prediction	79.29%	81.94%	79.93%

4. KESIMPULAN

Setelah melakukan eksperimen dan memperoleh hasil untuk pengujian prediksi penyakit jantung kesimpulan yang di dapat daripada metode algoritma optimasi ensemble pada algoritma Bootstrap Agregating (Bagging) dan juga Adaptive Boosting (AdaBoost) bisa meningkatkan daripada performa klasifikasi algoritma base estimator Decision Tree. Berdasarkan hasil perbandingan Analisa dari tiga dataset penyakit jantung, algoritma C4.5 menghasilkan nilai akurasi sebesar Heart Disease Dataset 84,78%, Heart Disease Prediction 72,96%, dan Heart Failure Prediction 79,29%. Setelah diterapkannya ensemble pada algoritma base estimator didapatkan pada metode ensemble bagging adalah sebesar Heart Disease Dataset 90,14%, Heart Disease Prediction 74,81%, dan Heart Failure Prediction 81,94%. Lalu pada metode ensemble AdaBoost mendapatkan hasil akurasi pada data Heart Disease Dataset 88,20, Heart Disease Prediction 78,89%, dan Heart Failure Prediction 79,93%.

UCAPAN TERIMA KASIH

Terima Kasih kepada dosen pembimbing saya yang selalu mendukung dan mengarahkan saya dalam penulisan jurnal ini, serta pihak-pihak kampus dan teman-teman yang selalu menyemangati dalam penulisan ini.

REFERENCES

- [1] Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, et al. Heart disease and stroke statistics. 2021 update: a report from the American Heart Associationexternal icon. *Circulation*. 2021;143:e254–e743.
- [2] enter for disease control and prevention. 2021, <https://www.cdc.gov/heartdisease/facts.htm>. Diakses pada 2 Januari 2022 pukul 23.12.
- [3] Cheryl D. Fryar, M.S.P.H.; Te-Ching Chen, Ph.D.; and Xianfen Li, M.S. Prevalence of Uncontrolled Risk Factors for Cardiovascular Disease: United States, 1999–2010, NCHS Data Brief, No. 103, August 2012.



- [4] A. Chauhan, A. Jain, P. Sharma, V. Deep, prediksi penyakit jantung menggunakan pembelajaran aturan evolusioner, dalam: Konferensi Internasional ke-4 2018 Kecerdasan Komputasi & Teknologi Komunikasi (CICT), Ghaziabad ,2018, hlm. 1-4.
- [5] Larose, Daniel T. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons, Inc.
- [6] Dai Qin-yun., Zang Chun-Ping., Wu Hao. 2016. Research of Decision tree Classification Algorithm in Data Mining. Dept. of Electric and Electronic Engineering, Shijiazhuang Vocational and Technology Institute. China
- [7] Sharma, R., Purushottam, Saxena, K. 2016. Efficient Heart Disease Prediction System using Decision Tree. International Conference on Computing, Communication and Automation (ICCCA), Noida, India, 15-16 May. 72-77. DOI: 10.1109/CCAA.2015.7148346
- [8] Zhang, S. (2012). Decision tree classifiers sensitive to heterogeneous costs. *Journal of Systems and Software*, 85(4), 771–779. doi:10.1016/j.jss.2011.10.007
- [9] Hou, S., Hou, R., Shi, X., Wang, J., & Yuan, C. 2014. Research on C5.0 Algorithm Improvement and the Test in Lightning Disaster Statistics. *International Journal of Control and Automation*, 7(1), 181-190.
- [10] Kavitha, R., Kannan, E. 2016. An Efficient Framework for Heart Disease Classification using Feature Extraction and Feature Selection Technique in Data Mining. *International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pp. 1-5.
- [11] Pristyanto, Yoga. 2019. “Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset.” *Jurnal TEKNOINFO* 13 (1): 11–16.
- [12] Puspitawuri, Annisa, Edy Santoso, and Candra Dewi. 2019. “Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor Dan Naïve Bayes.” *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 3 (4): 3319–24. <https://j-ptiik.ub.ac.id/index.php/jptiik/article/view/4916>.
- [13] Adelina, Vina, Dian Eka Ratnawati, and M. Ali Fauzi. 2018. “Klasifikasi Tingkat Risiko Penyakit Stroke Menggunakan Metode GA-Fuzzy.” *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer* 2 (September): 3015–21. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2513>.
- [14] Amini, Leila, Reza Azarpazhouh, Mohammad Taghi Farzadfar, Sayed Ali Mousavi, and Farahnaz Jazaieri. 2013. “Prediction and Control of Stroke by Data Mining.” *International Journal of Preventive Medicine* 4 (May): 245–49.
- [15] Bisri, Achmad. 2015. “Penerapan Adaboost Untuk Penyelesaian Ketidakseimbangan Kelas Pada Penentuan Kelulusan Mahasiswa Dengan Metode Decision Tree.” *Journal of Intelligent System* 1 (1):27-32.
- [16] Kansadub, Teerapat, and Sotarath Thammaboosadee. 2015. “Stroke Risk Prediction Model Based on Demographic Data.” 2015 8th Biomedical Engineering International Conference (BMEiCON),3-5.
- [17] Byna, Agus, and Muhammad Basit. 2020. “Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes.” *Jurnal Sisfokom (Sistem Informasi Dan Komputer)* 09 (November): 407–11.
- [18] D.Senthil Kumar , G.Sathyadevi and S.Sivanesh. (2016). Decision Support System for Medical Diagnosis Using Data Mining. *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 3, No. 1.
- [19] Tayeb, Shahab., Pirouz, Matin., Sun, Johann., Hall, Kaylee., Chang, Andrew., Li, Jessica., Song, Connor., Chauhan, Apoorva., Ferrá, Michael., Sager, Theresa., Zhan, Justin & Latifi, Shahram. 2018. Toward Predicting Medical Conditions Using K-Nearest Neighbors. *IEEE International Conference on Big Data* 3897– 3903.
- [20] Muslim, Much Aziz, Aldi Nurzahputra, and Budi Prasetyo. 2018. “Improving Accuracy of C4 . 5 Algorithm Using Split Feature Reduction Model and Bagging Ensemble for Credit Card Risk Prediction.” 2018 International Conference on Information and Communications Technology (ICOIAC), no.1996: 141–45.
- [21] Hana Rasheid Esmaeel (2020). Analysis of classification learning algorithms. *Indonesian Journal of Electrical Engineering and Computer Science* Vol. 17, No. 2, February 2020, pp. 1029~1039
- [22] Hand, David J. 2007. *Principles of Data Mining*. Springer:London.
- [23] Gorunescu, F. 2011. *Data Mining Concepts, Models, and Techniques*. Springer: Berlin.
- [24] Alpaydin, E. 2016. *Machine Learning: the New AI*. MIT Press.
- [25] Carlos J. Mantas, Joaquín Abellán, Javier G. Castellano (2016). Analysis of Credal-C4.5 for classification in noisy domains. *Express System With Applications* 61 314-326 Elsevier.
- [26] Latha, C Beulah Christalin, and S Carolin Jeeva. 2019. “Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques.” *Informatics in Medicine Unlocked* 16 (June): 100203. <https://doi.org/10.1016/j.imu.2019.100203>.
- [27] Tran, Cao Truong, Mengjie Zhang, Peter Andreae, and Bing Xue. 2017a. “Bagging and Feature Selection for Classification with Incomplete Data.” *European Conference on the Applications of Evolutionary* 10199 (April): 471–86.
- [28] Schapire, R. E., & Freund, Y. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55, 119-139. <https://doi.org/10.1088/00344885/55/7/004>.
- [29] Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*. 33, 1-39. <https://doi.org/10.1007/s10462-009-9124-7>.
- [30] Ramdhani, Lis Saumi. 2016. “Penerapan Particle Swarm Optimization (PSO) Untuk Seleksi Atribut Dalam Meningkatkan Akurasi Prediksi Diagnosis Penyakit Hepatitis Dengan Metode Algoritma C4 . 5.” *Jurnal Ilmu Komputer, Manajemen Dan Sosial (SWABUMI)* IV (1): 1-5.