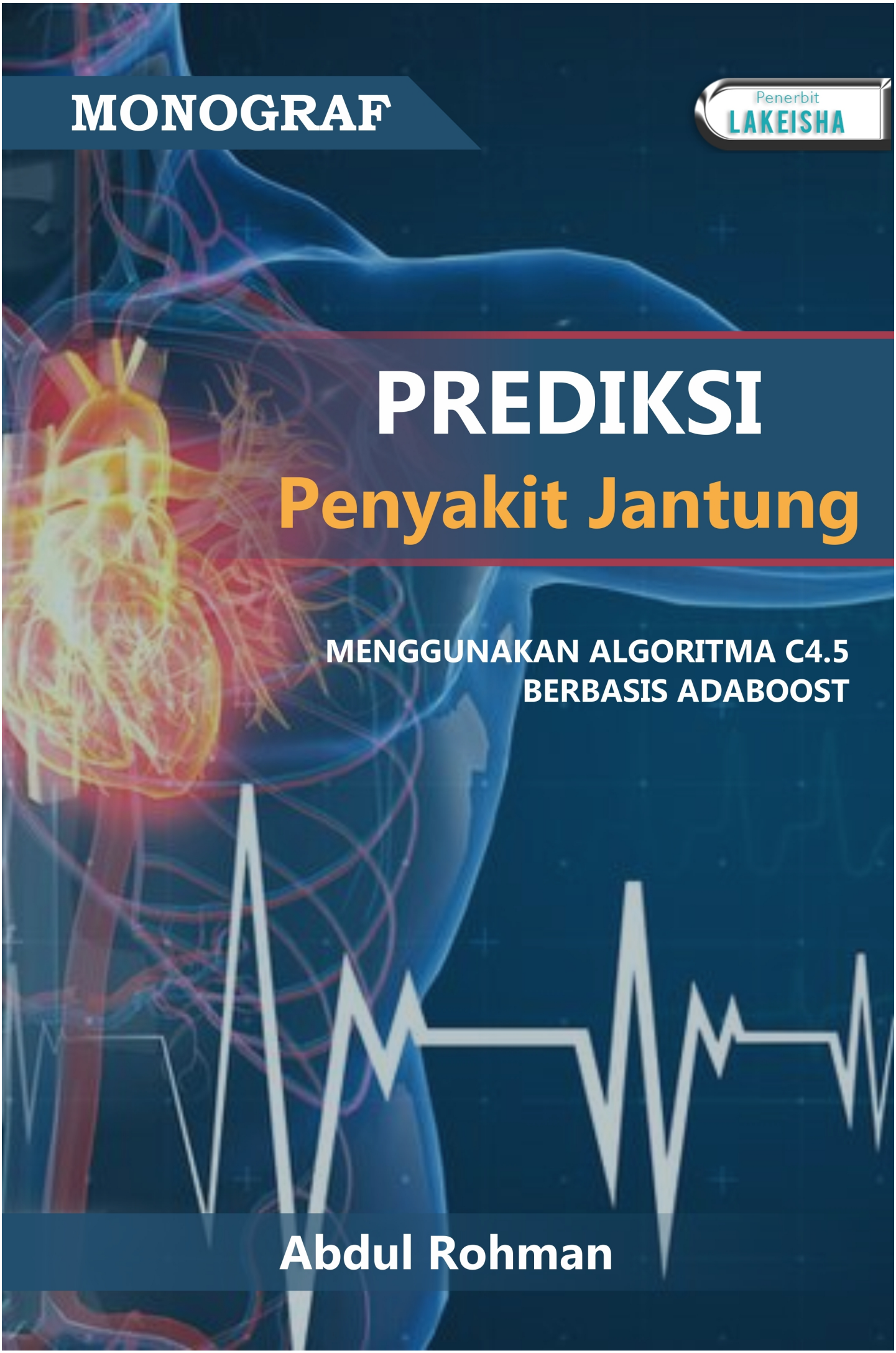
MONOGRAF

Penerbit
LAKEISHA



PREDIKSI
Penyakit Jantung

MENGGUNAKAN ALGORITMA C4.5 BERBASIS ADABOOST

Abdul Rohman

MONOGRAF

PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN

ALGORITMA C4.5 BERBASIS *ADABOOST*

Abdul Rohman

MONOGRAF

PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA C4.5 BERBASIS *ADABOOST*

Penerbit Lakeisha

2021

MONOGRAF

PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA C4.5 BERBASIS *ADABOOST*

Penulis:
Abdul Rohman

Editor: Dewi Kusumaningsih
Layout: Yusuf Deni Kristanto
Desain Cover: Tim Lakeisha
Cetak I November 2021
15.5 cm × 23 cm, 78 Halaman
ISBN: 978-623-420-004-1

Diterbitkan oleh Penerbit Lakeisha
(**Anggota IKAPI No.181/JTE/2019**)

Redaksi
Srikaton, RT 003, RW 001, Pucangmiliran,
Tulung, Klaten, Jawa Tengah
Hp. 08989880852, Email: penerbit_lakeisha@yahoo.com
Website: www.penerbitlakeisha.com

KATA PENGANTAR

Puji Syukur Kepada Allah yang Maha Pengasih dan Penyayang atas segala limpahanNya sehingga buku monograf "Prediksi Penyakit Jantung Menggunakan Algoritma *C4.5* Berbasis *Adaboost*" ini dapat terselesaikan dengan baik.

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di Indonesia, penyakit jantung sangat berbahaya karena dapat menimbulkan serangan jantung dan kematian mendadak. Oleh karena perlu kita harus waspada terhadap gejala-gejalanya.

Dengan adanya suatu prediksi terhadap penyakit jantung yang tepat, maka dapat dijadikan suatu informasi untuk mengenali, memahami dan mewaspadai terjadi serangan jantung. Atribut yang digunakan sebagai acuan timbulnya gejala yaitu; umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, kecepatan detak jantung, angina induksi, oldpeak, sengemt_st, floirosopy dan denyut jantung.

Algoritma *Decision Tree C4.5* berbasis *Adaboost* merupakan metode data mining yang baik dan banyak digunakan untuk membuat suatu prediksi terutama penyakit jantung, karena

mudah untuk diphami melalui hasil decision tree atau pohon keputusannya dan juga nilai akurasi performancenya yang baik.

Dengan permasalahan-permasalahan diatas, maka penulis ingin berbagi informasi terkait "prediksi penyakit jantung dengan menggunakan algoritma *C4.5* berbasis *Adaboost"* dalam bentuk buku monograf atau hasil penelitian.

Penulis Menyadari bahwa buku monograf ini masih jauh dari kata sempurna. Untuk itu, segala masukan sangat kami harapkan. Semoga buku ini bisa memberikan manfaat untuk Kita semua. Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu dalm penerbitan buku ini.

Ungaran, Nopember 2021

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Industri kesehatan memiliki sejumlah besar data kesehatan, namun sebagian besar data tersebut tidak diolah untuk mengetahui informasi tersembunyi untuk dijadikan pengambilan keputusan yang efektif oleh para praktisi kesehatan. Pengambilan keputusan atas dasar data dan informasi yang akurat akan menghasilkan keputusan dan prediksi penyakit menjadi tepat sasaran.

Jantung merupakan suatu organ otot berongga yang terletak dipusat dada yang menompa darah lewat pembuluh darah. Penyakit jantung terjadi karena penyumbatan sebagian atau total dari suatu lebih pembuluh darah, akibat dari adanya penyumbatan maka dengan sendirinya suplai energi kimiawi ke otot jantung berkurang, sehingga terjadi gangguan keseimbangan antara suplai dan kebutuhan darah (Rohman 2017). Penyakit jantung di Indonesia merupakan penyakit nomor satu yang mendorong angka kematian yang cukup tinggi, sehingga sampai sekarang penyakit tersebut ditakuti oleh manusia.

Faktor gejala yang terdiagnosa sebagai penyakit jantung diantaranya adalah jenis sakit dada (chest pain),tekanan darah tinggi (tresbps), kolesterol (chol), nilai tes EKG (resting electrodiagraphic(restecg)), denyut jantung (thalach) dan kadar gula (fasting blood sugar(FBS) (Aulia 2018). Dan beberapa factor

lainnya yang mengindentifikasi bahwa sesorang mempunyai penyakit jantung.

Penyakit jantung mempunyai beberapa jenis meliputi: penyakit jantung koroner, angina, serangan jantung dan gagal jantung (Utomo, Sirait, and Yunis 2020). Menurut The World Heart Federation, penyakit jantung adalah penyakit cardiovarsular yang sering menjangkit pada anak-anak dan orang dewasa (Salvi 2016). Dengan demikian menjadikan perhatian dalam masalah kesehatan terutama di negara-negara berkembang.

WHO (World Heart Organization) memperkirakan bahwa 12 juta kematian terjadi seluruh dunia, setiap tahun karena penyakit jantung (Annisa 2019). Diseluruh Amerika kematian akibat penyakit jantung mencapai 925.227 penderita, yakni dari seluruh kematian atau setiap hari 2600 penduduk meninggal akibat jantung (Salvi 2016).

Menurut Palaniappan dan Awang, penyakit jantung perlu diprediksi karena keputusan klinis sering kali dibuat oleh dokter berdasarkan intuisi dan bukan pengalaman pengetahuan yang didapat atas data yang tersembunyi dalam database (K, M, and R 2018). Dengan keputusan klinis yang buruk akan menyebabkan ketidaktepatan diagnose, biaya medis yang berlebihan sehingga mempengaruhi kualitas pelayanan dan perawatan pada pasien.

Data mining adalah menganalisis pengamatan dataset untuk menemukan hubungan yang tak terduga dan untuk meringkas data dengan cara baru yang baik dimengerti dan bermanfaat bagi pemilik data (Olusola, Oladele, and Abosede 2016). Sedangkan Witten (Parlar and Acaravci 2017), berpendapat bahwa teknik yang digunakan dalam data mining untuk menemukan dan menggambarkan pola struktur dalam data, sebagai alat untuk membantu menjelaskan data dan membuat prediksi dari data tersebut.

Banyak penelitian prediksi penyakit jantung dengan teknik klasifikasi Data mining, diantaranya dilakukan oleh Palaniappan dan Awang (K, M, and R 2018). dengan melakukan komporasi 3 metode yaitu Naives Bayes, Decision Tree, dan Artificial Neural Network (ANN) dengan total kasus 909 dan 15 atribut. Hasil dari penelitian tersebut metode Decision tree menghasilkan nilai terbaik.

Penelitian yang dilakukan oleh Anbarasi dkk [8] dalam memprediksi kelangsungan hidup penyakit jantung dengan berdasarkan 909 kasus dan 6 Atribut dengan menggunakan metode Naïve Bayes, Decision Tree dan Clasification Via Clustering. Hasil penelitian tersebut metode Decision Tree menghasilkan nilai terbaik.

Selain itu juga S.B. Kotsiantis dalam review papernya menjelaskan, bahwa metode Decision Tree mempunyai kelebihan-kelebihan dalam mengolah dataset penyakit jantung yaitu dari segi; kecepatan dalam klasifikasi, tiap atribut bersifat diskrit, binari dan kontinue, serta transparansi pengetahuan atau klasifikasi (Soysal and Schmidt 2010).

Metode adaboost adalah teknik optimasi yang handal dengan mengkombinasikan beberapa pengklasifikasian dasar (multiple base classifiers) untuk menghasilkan suatu pengklasifikasian yang kuat (strong classifier) (Bisri and Wahono 2015).

Berdasarkan atas penelitian diatas, peneliti akan memilih metode Decision tree dalam memprediksi penyakit jantung. Dalam penelitian ini akan dilakukan penerapan algoritma Decision tree (C4.5) menggunakan metode adaboost dengan mengoptimal atribut-atribut yang berasal dari dataset yang terpercaya untuk memprediksi penyakit jantung dengan tujuan agar akurasi menjadi meningkat.

1.2 Permasalahan

Berdasarkan latar belakang diatas algoritma C4.5 banyak digunakan pada penelitian diberbagai bidang, khususnya dalam data mining. Banyak komparasi algoritma yang dilakukan peneliti-peneliti sebelumnya, algoritma C4.5 dalam prediksi penyakit jantung cenderung memiliki nilai terbaik diantara algoritma lainnya, akan tetapi dataset sangat berpengaruh terhadap akurasi. Dalam penelitian ini dapat disimpulkan bahwa tingkat akurasi algoritma C4.5 masih belum mencapai level excellence.

Maka permasalahan dalam penelitian ini yaitu seberapa akurat model C4.5 yang ditingkatkan dengan metode adaboost dalam memecahkan masalah prediksi jantung

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah melakukan optimasi algoritma C4.5 berbasis adaboost dengan melakukan perulangan (iteration) dan attribute wighting untuk meningkat akurasi dalam prediksi penyakit jantung.

Manfaar penelitian ini adalah:

a. Manfaat praktis dari dari penelitian ini adalah diharapkan agar dapat digunakan oleh para praktisi kesehatan seperti dokter untuk sebagai alat bantu untuk prediksi penyakit jantung.
b. Manfaat teoritis dari hasil penelitian ini diharapkan memberikan sumbangsih bagi pengembang teori yang berkaitan dengan prediksi penyakit jantung dengan menggunakan algoritma C4.5 berbasis metode adaboost.

BAB II
TINJUAN PUSTAKA

2.1 Penelitian Terkait

Penelitian tentang prediksi penyakit jantung cukup banyak. Berikut ini beberapa penelitian terkait tentang prediksi penyakit jantung yaitu sebagai berikut:

Penelitian yang dilakukan oleh Sellappan Palaniappan dan Rafiah Awang tahun 2008 melakukan komporasi 3 metode yaitu Naives Bayes, Decision Tree, dan Artificial Neural Network (ANN) dengan total kasus 909 dan 15 atribut. Hasil dari penelitian tersebut metode Decision tree menghasilkan nilai terbaik yaitu 89%, Naïve Bayes 86,53% dan Artificial Neural Network (ANN) 85,53% (K, M, and R 2018).

Penelitian yang dilakukan oleh M. Anbarasi, E. Anupriya dan Inyenar dalam memprediksi kelangsungan hidup penyakit jantung dengan berdasarkan 909 kasus dan 6 Atribut dengan menggunakan metode Naïve Bayes, Decision Tree dan Clasification Via Clustering. Hasil penelitian tersebut metode Decision Tree menghasilkan nilai terbaik yaitu 99,2%, Naïve Bayes 96,5% dan Clasication Via Clustering 88,3% (M Anbarasi, E Anupriya 2010).

Selain itu juga S.B. Kotsiantis dalam review papernya menjelaskan, bahwa metode Decision Tree mempunyai kelebihan-kelebihan dalam mengolah dataset penyakit jantung yaitu dari segi; kecepatan dalam klasifikasi, tiap atribut bersifat diskrit, binari dan kontinue, serta transparansi pengetahuan atau klasifikasi (Soysal and Schmidt 2010).

Dalam penelitian ini akan menggunakan algoritma C4.5 berbasis metode adaboost dengan cara meningkatkan atribut (attribute weigthing), terhadap data penyakit jantung sehingga dapat meningkatkan tingkat akurasinya.

2.1 Landasan Teori

2.1.1 Penyakit Jantung

Jantung adalah organ berupa otot, berbentuk kerucut, berongga dengan basisnya diatas dan puncaknya dibawah. Yang fungsinya untuk memompa bersih ke seluruh tubuh dan darah kotor ke paru-paru. Jika terjadi gangguan pada jantung maka fungsi pemompaan darah akan terganggu bahkan bisa mengakibatkan kematian.

Penyakit jantung adalah penyempitan atau penyumbatan (*arteriosclerosis*) pembuluh arteri koroner yang disebabkan oleh penumpukan dari zat-zat lemak (*kolestrol, trigliserida*) yang makin lama makin banyak dan menumpuk dibawah lapisan terdalam (*endotelium*) dari dinding pembuluh nadi (Iskandar, Hadi, and Alfridsyah 2017).

Berdasarkan dataset penyakit jantung di UCI (Univercity of California Irvine) terdapat 14 atribut yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elekrokardiografi, tekanan darah, angina induksi, oldpeak, segmen_st, flaurosopy, denyut jantung dan hasil sebagai label yang terdiri atas *healthy* (sehat) dan *sick* (sakit). Semua atribut tersebut selain hasil merupakan hal-hal yang mempengaruhi terjadinya penyakit jantung.

1. Umur

   Pada survei rumah tangga mengenai kesehatan yang telah dilakukan oleh Badan Litbang Depkes RI, penyakit kardiovakuler angka prevalensinya bergeser dari urutan ke-9

pada tahun 1972, menjadi urutan ke-6 pada tahun 1980 dengan 5,9 kasus per 1000 penduduk. Secara spesifik prevalensi penyakit kardiovaskuler khususnya *infarct myocard* pada kelompok umur kurang dari 40 tahun sebesar 3,1 % dan pada kelompok umur 40 s.d 49 tahun sebesar 19,9 %.(19) Sedangkan insiden serupa yang terjadi di Jawa Tengah, kejadian *infarct myocard* secara umum sebesar 1,03 % dan gejala *angina pectoris* (nyeri ulu hati) sebesar 0,50 % (Amelia. 2015)

2. Jenis kelamin

   Laki-laki memiliki risiko lebih besar terkena serangan jantung dan kejadiannya lebih awal dari pada wanita. Morbiditas penyakit jantung pada laki-laki dua kali lebih besar dibandingkan dengan wanita dan kondisi ini terjadi hampir 10 tahun lebih dini pada laki-laki darpada perempuan.

3. Jenis sakit dada (angina)

   Angina adalah tipe nyeri pada dada yang disebabkan berkurangnya darah yang mengalir pada otot jantung. Angina adalah gejala dari *coronary artery disease* (*CAD*) atau penyakit coroner. Pada penyakit ini otot jantung tidak mendapat cukup darah yang kaya akan oksigen. Berdasarkan dataset UCI jenis sakit dada terdiridari: angina khas, angina atipikal, pectoris dan asimtomatik

4. Tekanan darah (hipertensi)

   Peningkatan tekanan darah meningkatkan resistensi terhadap pemompaan darah dari ventrikel kiri, sebagai akibatnya terjadi hipertropi ventrikel untuk meningkatkan kekuatan kontraksi. Kebutuhan oksigen oleh miokardium akan meningkat akibat hipertrofi ventrikel, hal ini mengakibat peningkatan beban kerja jantung yang pada akhirnya menyebabkan angina dan infark miokardium. Tekanan darah yang di ukur adalah ketika sedang istirahat.

5. Kolestrol.

   Pada latihan fisik akan terjadi dua perubahan pada sistem kardiovaskuler, yaitu peningkatan curah jantung dan redistribusi aliran darah dari organ yang kurang aktif ke organ yang aktif. Aktivitas aerobik secara teratur menurunkan risiko penyakit jantung. Disimpulkan juga bahwa olah raga secara teratur akan menurunkan tekanan darah sistolik, menurunkan kadar katekolamin di sirkulasi, menurunkan kadar kolesterol dan lemak darah, meningkatkan kadar HDL lipoprotein, memperbaiki sirkulasi koroner dan meningkatkan percaya diri.

6. Kadar gula

   Kadar gula mempengaruhi terjadinya penyakit jantung. Kecenderungan kadar gula darah puasa > 126 mg/dl merupakan faktor risiko untuk terjadinya penyakit jantung.

7. EKG (Elektrokardiografi)

   Secara umum penderita penyakit jantung biasanya tampak cemas, gelisah, pucat dan berkeringat dingin. Denyut nadi umumnya cepat (takhikardi), irama tidak teratur, tetapi dapat pula denyut nadi lambat (bradikardia). Hipertensi maupun hipotensi dapat terjadi pada penderita ini. Meskipun kadang-kadang kurang jelas pada pemeriksaan fisik pada jantung, tetapi pemeriksaan menggunakan EKG (*elektrokardiografi*) akan sangat membantu memberi informasi.

8. Kecepatan Detak jantung (tekanan jantung)
   Peningkatan tekanan jantung mempengaruhi terjadinya serangan jantung. Saat stress dengan istirahat memperlihatkan kadar tekanan jantung yang berbeda-beda, kecenderungan orang yang sering stress akan menimbulkan peningkatan terhadap tekanan jantung,

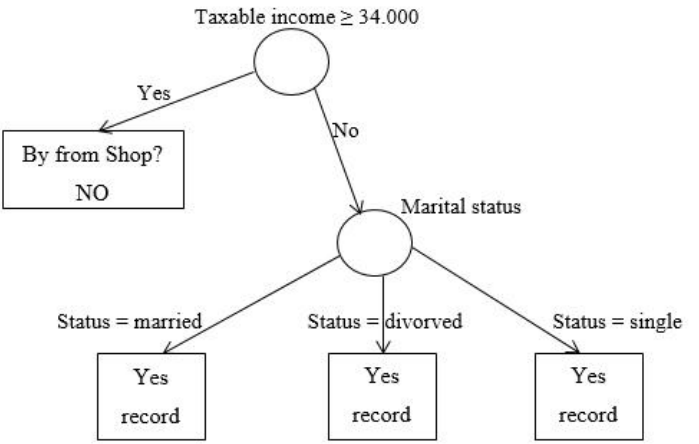
9. Angina induksi
   Latihan angina induksi merupakan faktor yang mempengaruhi penyakit jantung. Berdasarkan dataset UCI terdapat dua pilihan yaitu ya dan tidak dalam melakukan latihan angina induksi.
10. Oldpeak
    Oldpeak berdasarkan UCI adalah ST depresi disebabkan oleh latihan relatif untuk beristirahat
11. Segmen_st
    Kemiringan dari puncak latihan dengan nilai; condong keatas, datar dan condong.
12. Flourosopy
    Banyaknya nadi utama (0-3) yang diwarnai oleh flourosopy
13. Denyut jantung
    Denyut jantung dikategorikan 3 yaitu; keadaan normal, cacat tetap dan cacat sementara.

2.1.2 Algoritma C4.5

Algoritma *decision tree* digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar [12]. Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.

Model *decision tree* yang sudah berkembang antara lain: *IDS*, *C4.5* dan *CART* (*classification and regression tree*) (Budi Santosa 2007). Algoritma *C4.5* mirip sebuah pohon dimana terdapat node internal (bukan daun) yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Pohon keputusan dengan mudah dapat dikonversi ke aturan klasifikasi. Secara umum keputusan

pengklasifikasi pohon memiliki akurasi yang baik, namun keberhasilan penggunaan tergantung pada data yang akan diolah.



Gambar 2.1 Contoh konsep pohon keputusan sederhana (Gorunescu 2011).

Pada Gambar 2.1 variabel target untuk pohon keputusan adalah membeli pada toko, dengan pengklasifikasian ya atau tidak. Variabel *predictor* adalah taxable income (<34.000 atau > 34.000), marital status (married, divorced, atau single). Simpul akar merupakan simpul keputusan, pengujiannya apakah taxable income < 34.000 atau > 34.000.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma *C4.5* (Gorunescu 2011) yaitu:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing atribut atau berdasarkan nilai *index entropy* terendah. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus:

$$Entropy(i) = \sum_{j=1}^{m} \mathrm{f}(i,j).2\ \mathrm{f}[(i,j)] \qquad (2.1)$$

Keterangan:

i = himpunan kasus

m = jumlah partisi i

f(i,j) = proposi j terhadap i

3. Hitung nilai *gain* dengan rumus:

$$\text{gain} = -\sum_{i=1}^{p} \frac{ni}{n}.IE(i) \quad (2.2)$$

Keterangan:

p = jumlah partisi atribut

ni = proporsi ni terhadap i

n = jumlah kasus dalam n

4. Untuk menghitung gain ratio perlu diketahui suatu term baru yang disebut Split Information dengan rumus:

$$SplitInformation = -\sum_{t=1}^{c} \frac{S_1}{S} log_2 \frac{S_1}{S} \quad (2.3)$$

S_1 sampai S_c = c subset yang dihasilkan dari pemecahan S dengan menggunakan atribut A yang mempunyai sebanyak c nilai

5. Selanjutnya menghitung gain ratio

$$Gainratio(S,A) = \frac{Gain(S,A)}{SplitInformation(S,A)} \quad (2.4)$$

6. Ulangi langkah ke-2 hingga semua *record* terpartisi

Proses partisi pohon keputusan akan berhenti disaat:

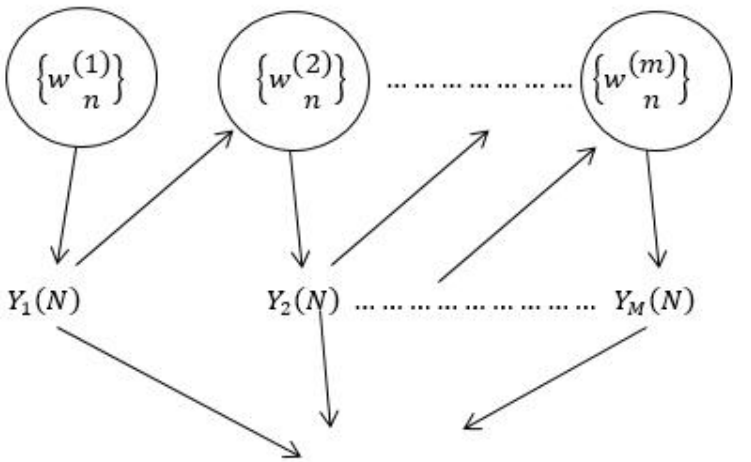
a. Semua tupel dalam *record* dalam simpul m mendapat kelas yang sama
b. Tidak ada atribut dalam *record* yang dipartisi lagi
c. Tidak ada *record* didalam cabang yang kosong.

2.1.3 Metode Adaboost

Pada dasarnya, metode boosting juga dapat meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara membangkitkan kombinasi dari suatu model, tetapi

hasil klasifikasi atau prediksi yang dipilih adalah model yang memiliki nilai bobot paling besar. Jadi, setiap model yang dibangkitkan memiliki atribut berupa nilai bobot. Salah satu algoritma boosting yang populer adalah *Adaboost*. *Adaboost* atau *adaptive boosting* adalah algoritma pembelajaran, pertama kali oleh Freund dan Scaphire tahun 1995.

Pada permasalahan klasifikasi dua kelas, dimana input vector adalah x1, x2,…., xN dan mempunyai target t1, t2,……. tN dimana $tn \in \{-1,1\}$. Setiap data diberikan parameter bobot (nilai awal adalah 1/N untuk semua data). Pada proses pelatihan pengklasifikasi dasar menggunakan bobot data untuk fungsi y(x) $\in \{-1,1\}$. Setiap tahapan algoritma, adaboost melatih klasifikasi baru menggunakan dataset dengan koefesien bobot yang diatur berdasarkan performansi hasil pelatihan pengklasifikasi sebelumnya, sehingga memberikan bobot besar data yang salah klasifikasi. Setelah pengklasifikasi dasar sudah melakukan proses pelatihan, selanjutnya dikombinasikan ke bentuk *committee* menggunakan koefisien dengan bobot berbeda ke pengklasifikasi dasar yang berbeda.



$$Y_M(N) = sign\left(\sum_m^M a_m\, y_m\,(X)\right) \qquad (2.5)$$

Gambar 2. 2 Ilustrasi skema metode adaboost

Setiap klasifikasi dasar $y_n(x)$ dilatih menggunakan bobot $W_n^{(m)}$ Tergantung pada performansi pengklasifikasi dasar sebelumnya $Y_{m-1}\,(x)$. Hasil pelatihan semua pengklasifikasi dasar dikombinasikan ke pengklasifikasi tertentu $y_m(x)$.

Metode adaboost adalah sebagai berikut:

1. Inisialisasi bobot data $\{W_n\}$ dengan $W_n^{(m)}$ untuk n = 1, 2, ……..,N.
2. For m = 1 ……..,M.
   a. Training $y_m(x)$ dengan meminimalkan fungsi kesalahan (*error function*) sebagai berikut:

$$J_m = \sum_{n=1}^{N} W^{(m)}{}_n I\,(y_m(x_n) \neq t_n \qquad (2.6)$$

Dimana: $I(y_m(x_n) \neq t_n$ adalah fungsi indikator sama dengan satu jika $(y_m(x_n) \neq t_n$ dan 0 lainnya.

b. Evaluasi kesalahan

$$\varepsilon_m = \frac{\sum_{n=1}^{N} w_n^{(m)} (y_m(x_n) \neq t_n}{\sum_{n=1}^{N} w_n^{(m)}} \qquad (2.7)$$

Dan kemudian digunakan evaluasi

$$a_m = In \left\{ \frac{1 - \varepsilon_m}{\varepsilon_m} \right\} \qquad (2.8)$$

c. Memperbaiki (update) bobot data

$$w_n^{(m+1)} = w_n^{(m)} \exp(a_m I(y_m(x_n) \neq t_n)\} \qquad (2.9)$$

3. Membuat prediksi menggunakan model terakhir sebagai berikut

$$Y_m(x) = sign\ (\sum_{m=1}^{M} a_m y_m\ (x)) \qquad (2.10)$$

2.1.4 Algoritma C4.5 berbasis Adaboost

Setelah melakukan tahapan dalam membuat sebuah pohon keputusan dengan algoritma *C4.5*, dilakukan pemberian bobot pada pohon tunggal sehingga menghasilkan hipotesa baru dan sebuah pohon keputusan baru dengan langkah-langkah sebagai berikut (Suwondo, Asmarajati, and Surahman 2013):

1. Menentukan bobot awal Wi = 1/n, dimana n = pengamatan
2. Penentuan iterasi dimana nomor iterasi m = 1,2,….M. Kemudian lakukan proses:
   a. Susun pohon tunggal dengan memperhatikan bobot Wi

b. Hitung tingkat kesalahan klasifikasi
c. Hitung α_m (evaluasi)
d. Menentukan bobot baru untuk setiap pengamatan

Tabel 2.1 Perhitungan bobot nilai iterasi 1…..m

Hipotesa	Error (em)	αm (evaluasi)	Wm	Hasil Hipotesa
1	0	0	1	0
(prediksi Healthy atau Sick)	(rumus 2.7)	(rumus 2.8)	(rumus 2.9)	(bobot baru)
58	0	0	1	0

3. Menentukan nilai akhir class
   Hasil αm dari masing-masing hipotesa dari iterasi 1 sampai 10 digabungkan dan didapatkan nilai + 1.

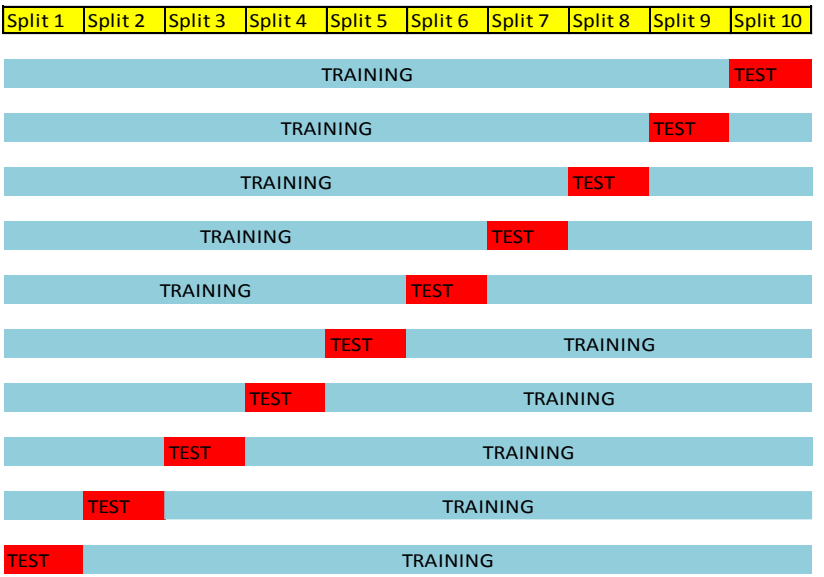
Tabel 2.2 Menentukan nilai akhir class

Hipotesa	αm (iterasi 1 sampai 10)	Sign αm1 + αm 2...10
(hipotesa terpilih)	(error > 0)	(rumus 2.8)

2.1.5 Pengujian K-Fold Cross Validation

Cross Validation adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Pratiwi and Ulama 2016). Dengan

menggunakan *cross validation* akan dilakukan percobaan sebanyak k. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau 10-*fold Cross Validation*.



Gambar 2.3 Ilustrasi 10 *Fold Cross Validation*

Pada gambar diatas terlihat bahwa tiap percobaan akan menggunakan satu data *testing* dan k-1 bagian akan menjadi data *training*, kemudian data *testing* itu akan ditukar dengan satu buah data *training* sehingga untuk tiap percobaan akan didapatkan data *testing* yang berbeda-beda.

2.1.6 Evaluasi dan Validasi

Dalam melakukan evaluasi pada algoritma *C4.5* dan algoritma *C4.5* optimasi dengan teknik *adaboost* dilakukan dengan

menggunakan model *counfusion matrix,* dan ROC *curve* (*Receiver Operating Characteristic*).

1. *Confusion Matrix*

Confusion matrix memberikan keputusan yang diperoleh dalam *traning* dan *testing* (Vulandari 2017), *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu 2011).

Tabel 2.3 Model *Confusion Matrix*

Classification	**Predicted class**		
Observed class		Class = yes	Class = no
	Class = yes	A (*true positive-tp*)	B (*false negative – fn*)
	Class = no	C (*false positive – fp*)	D (*true negative – tn*)

Keterangan:

True Positive (tp) = proporsi positif dalam data set yang diklasifikasikan positif

True Negative (tn)= proporsi negative dalam data set yang diklasifikasikan negative

False Positive (fp) = proporsi negatif dalam data set yang diklasifikasikan potitif

False Negative (fn)= proporsi negative dalam data set yang diklasifikasikan negatif

Berikut adalah persamaan model *confusion matrix*:

a. Nilai akurasi (acc) adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$acc = \frac{tp+tn}{tp+tn+fp+fn} \qquad (2.8)$$

b. Sensitivity digunakan untuk membandingkan proporsi tp terhadap tupel yang positif, yang dihitung dengan menggunakan persamaan:

$$sencitivity = \frac{tp}{tp+fn} \qquad (2.9)$$

c. Specificity digunakan untuk membandingan proporsi tn terhadap tupel yang negatif, yang dihitung dengan menggunakan persamaan:

$$sencitivity = \frac{tn}{tn+fp} \qquad (2.10)$$

d. PPV (*positive predictive value*) adalah proporsi kasus dengan hasil diagnosa positif, yang dihitung dengan menggunakan persamaan:

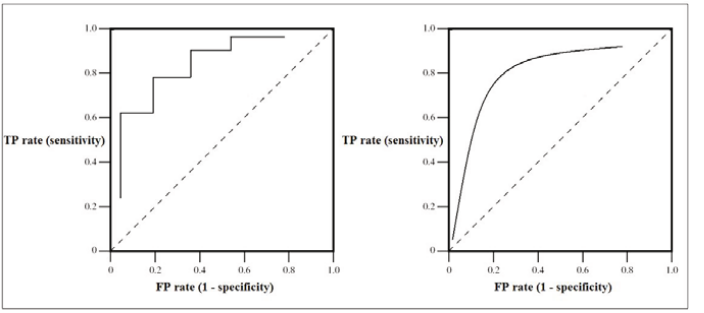
$$ppv = \frac{tp}{tp+fp} \qquad (2.11)$$

e. NPV (*negative predictive value*) adalah proporsi kasus dengan hasil diagnosa negatif, yang dihitung dengan menggunakan persamaan:

$$npv = \frac{tn}{tn+fn} \qquad (2.12)$$

2. *Curve ROC*

Untuk dapat melihat akurasi secara manual dilakukan perbandingan klasifikasi menggunakan c*urva ROC* hasil eksperesi dari *confusion matrix*. Kurva ROC (*Receiver Operating Characteristic)* adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual [16]. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (fp) pada sumbu *X* dan proporsi positif benar (tp) pada sumbu *Y*. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada (fp = 0) dan nilai positif benar adalah tinggi (tp = 1). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif {-1}, dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif {1}. Grafik ROC menggambarkan *trade-off* antara manfaat („*true positives*") dan biaya („*false positives*"). Berikut tampilan dua jenis kurva ROC (*discrete* dan *continous*).



Gambar 2.4 Grafik ROC (*discrete* dan *continous*) (Gorunescu 2011)

Pada Gambar 2.4 garis diagonal membagi ruang ROC, yaitu:

1. poin diatas garis diagonal merupakan hasil klasifikasi yang baik.
2. (b) point dibawah garis diagonal merupakan hasil klasifikasi yang buruk.

Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Tingkat akurasi dapat di diagnosa sebagai berikut [12]:

Akurasi 0.90 – 1.00 = *Excellent classification*

Akurasi 0.80 – 0.90 = *Good classification*

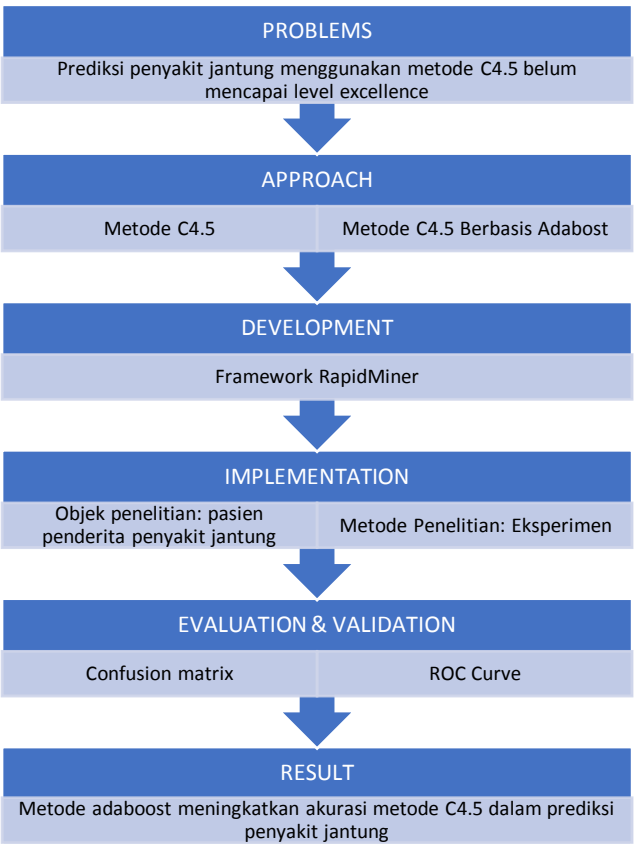
Akurasi 0.70 – 0.80 = *Fair classification*

Akurasi 0.60 – 0.70 = *Poor classification*

Akurasi 0.50 – 0.60 = *Failure*

2.3 Kerangka Pemikiran dan hipotesis

Kerangka pemikiran dalam proposal ini dimulai dari kurang sadarnya masyarakat atas gejala penyakit jantung serta kurang akuratnya penerapan algoritma C4.5 dalam memprediksi penyakit jantung.



Gambar 2.5 Kerangka Pemikiran

BAB III
METODOLOGI PENELITIAN

3.1 Desain Penelitian

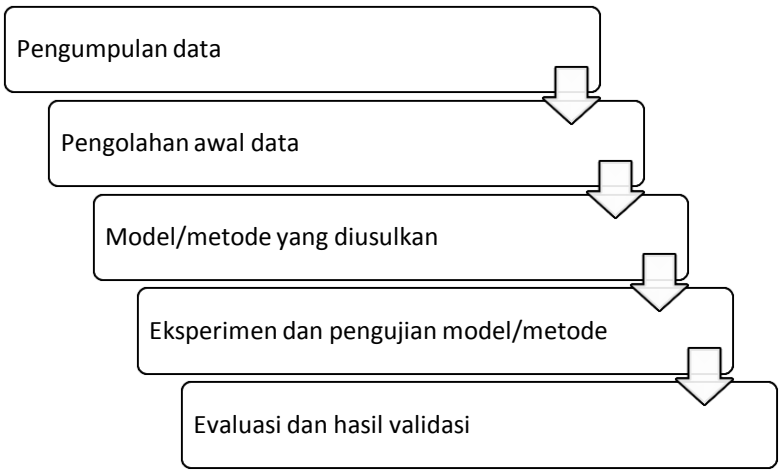
Dalam penelitian ini menggunakan data pasien yang melakukan pemeriksaaan penyakit jantung yang didapat dari UCI (*Universitas California, Invene*) *Machine Learning Repository* (Derisma 2020). Data pemeriksaaan penyakit jantung ini akan diolah menggunakan algoritma *C4.5* dan dengan menggunakan motede optimasi *adaboost* sehingga diperoleh metode yang akurat dan dapat digunakan sebagai aturan dalam prediksi penyakit jantung.

Penelitian ini adalah penelitian *experiment* yang melibatkan penyelidikan tentang perlakuan pada parameter dan variabel yang semuanya tergantung pada peneliti itu sendiri. *software* dan *hardware* sebagai alat bantu dalam penelitian ini adalah sebagai berikut:

Tabel 3.1 Spesifikasi hardware dan software

Software	Hardware
Sistem operasi: Windows XP SP III 32 bit **Data Mining: RapidMiner Versi 5**	CPU: Dual Core 1,7 Ghz Ram 2 GB, Hdd 160Gb

Dalam penelitian ini, ada beberpa langkah-langkah untuk melakukan proses penelitian sebagai berikut:



Gambar 3.1 Tahapan Penelitian

Pada tahapan penelitian dibagi menjadi 5 tahapan antara lain pengumpulan data, pengolahan awal data, model atau metode yang diusulkan atau dikembangkan, eksperimen dan pengujian model atau metode dan yang terakhir evaluasi dan validasi hasil.

3.2 Pengumpulan data

Dalam pengumpulan data peneliti mengambil data dari UCI adalah data pasien yang memeriksakan penyakit jantung dengan hasil yang didapat sebanyak 867 orang yang diperiksa dan sebanyak 364 pasien terdeteksi sakit, sehingga 503 pasien terdeteksi sehat (Derisma 2020). Dataset tersebut adalah penggabungan antara dataset dari Cleveland yang terdiri dari 303 pasien, data dari statlog yang terdiri dari 270 pasien, dan data dari hungaria terdiri dari 294 pasien.

Dengan atribut dari setiap penyakit jantung yang diperiksa adalah umur, jenis kelamin, jenis sakit dada, tekanan darah,

kolestrol, kadar gula, elektrokardiografi, tekanan darah, angina induksi, oldpeak, segmen_st, flaurosopy, denyut jantung dan hasil sebagai label yang terdiri atas *healthy* (sehat) dan *sick* (sakit). Data pasien yang memeriksakan penyakit jantung bisa di lihat pada tabel yaitu sebagai berikut:

Tabel 3.2 Atribut dan penyakit jantung

no	umur	jenis kelamin	jenis sakitdada	tekanan darah	kolestrol	kadar gula	elektrokardiografi	tekanan jantung	angina induksi	oldpeak	segmen_st	flaurosopy	denyut jantung	hasil
1	63.0	male	typ_angina	145.0	233.0	t	left_vent_hyper	150.0	no	2.3	down	0.0	fixed_defect	healthy
2	67.0	male	asympt	160.0	286.0	f	left_vent_hyper	108.0	yes	1.5	flat	3.0	normal	sick
3	67.0	male	asympt	120.0	229.0	f	left_vent_hyper	129.0	yes	2.6	flat	2.0	reversable_defect	sick
4	37.0	male	non_anginal	130.0	250.0	f	normal	187.0	no	3.5	down	0.0	normal	healthy
5	41.0	female	atyp_angina	130.0	204.0	f	left_vent_hyper	172.0	no	1.4	up	0.0	normal	healthy
6	56.0	male	atyp_angina	120.0	236.0	f	normal	178.0	no	0.8	up	0.0	normal	healthy
7	62.0	female	asympt	140.0	268.0	f	left_vent_hyper	160.0	no	3.6	down	2.0	normal	sick
8	57.0	female	asympt	120.0	354.0	f	normal	163.0	yes	0.6	up	0.0	normal	healthy
9	63.0	male	asympt	130.0	254.0	f	left_vent_hyper	147.0	no	1.4	flat	1.0	reversable_defect	sick
10	53.0	male	asympt	140.0	203.0	t	left_vent_hyper	155.0	yes	3.1	down	0.0	reversable_defect	sick
11	57.0	male	asympt	140.0	192.0	f	normal	148.0	no	0.4	flat	0.0	fixed_defect	healthy
12	56.0	female	atyp_angina	140.0	294.0	f	left_vent_hyper	153.0	no	1.3	flat	0.0	normal	healthy
13	56.0	male	non_anginal	130.0	256.0	t	left_vent_hyper	142.0	yes	0.6	flat	1.0	fixed_defect	sick
14	44.0	male	atyp_angina	120.0	263.0	f	normal	173.0	no	0.0	up	0.0	reversable_defect	healthy
15	52.0	male	non_anginal	172.0	199.0	t	normal	162.0	no	0.5	up	0.0	reversable_defect	healthy
16	57.0	male	non_anginal	150.0	168.0	f	normal	174.0	no	1.6	up	0.0	normal	healthy
17	48.0	male	atyp_angina	110.0	229.0	f	normal	168.0	no	1.0	down	0.0	reversable_defect	sick
18	54.0	male	asympt	140.0	239.0	f	normal	160.0	no	1.2	up	0.0	normal	healthy
19	48.0	female	non_anginal	130.0	275.0	f	normal	139.0	no	0.2	up	0.0	normal	healthy
20	49.0	male	atyp_angina	130.0	266.0	f	normal	171.0	no	0.6	up	0.0	normal	healthy

3.3 Pengolahan awal data

Data yang diperoleh dari UCI akan di *preprocessing* terlebih dahulu supaya data berkualitas dengan cara manual.

Teknik dalam *preprocessing* (Fadli 2011) yaitu:

a. *Data cleaning* bekerja membersihkan nilai kosong, tidak konsisten atau tupel kosong (*missing value* dan *noisy*).
b. *Data integration* menyatukan tempat penyimpanan (arsip) yang berbeda dalam satu arsip.
c. *Data reduction* jumlah atribut yang digunakan untuk data training terlalu besar sehingga ada beberapa atribut yang tidak diperlukan dihapus.

Dalam penelitian ini akan menggunakan teknik Data cleaning, data yang nilai atributnya kosong akan dihapus supaya data menjadi lebih valid dan berkualitas. Dalam penelitian jumlah record/data pasien adalah 867, setelah melakukan data cleaning maka dapat ditemukan data dengan atribut yang kosong berjumlah 300 dan data

terisi semua berjumlah 567. Jadi data yang diolah dan diteliti sebanyak 567 dengan keadaan sakit sejumlah 257 orang dan keadaan sehat sejumlah 310 orang. Data pasien yang memeriksakan penyakit jantung setalah *Data Cleaning* bisa di lihat pada tabel 3.3. yaitu sebagai berikut.

Tabel 3.3 Dataset setelah *Data Cleaning*

No	umur	jenis kelamin	jenis sakitdada	tekanan darah	kolestrol	kadar gula	elektrokardiografi	tekanan jantung	angina induksi	oldpeak	segmen_st	flaurosopy	denyut jantung	hasil
1	63.0	male	typ_angina	145.0	233.0	t	left_vent_hyper	150.0	no	2.3	down	0.0	fixed_defect	healthy
2	67.0	male	asympt	160.0	286.0	f	left_vent_hyper	108.0	yes	1.5	flat	3.0	normal	sick
3	67.0	male	asympt	120.0	229.0	f	left_vent_hyper	129.0	yes	2.6	flat	2.0	reversable_defect	sick
4	37.0	male	non_anginal	130.0	250.0	f	normal	187.0	no	3.5	down	0.0	normal	healthy
5	41.0	female	atyp_angina	130.0	204.0	f	left_vent_hyper	172.0	no	1.4	up	0.0	normal	healthy
6	56.0	male	atyp_angina	120.0	236.0	f	normal	178.0	no	0.8	up	0.0	normal	healthy
7	62.0	female	asympt	140.0	268.0	f	left_vent_hyper	160.0	no	3.6	down	2.0	normal	sick
8	57.0	female	asympt	120.0	354.0	f	normal	163.0	yes	0.6	up	0.0	normal	healthy
9	63.0	male	asympt	130.0	254.0	f	left_vent_hyper	147.0	no	1.4	flat	1.0	reversable_defect	sick
10	53.0	male	asympt	140.0	203.0	t	left_vent_hyper	155.0	yes	3.1	down	0.0	reversable_defect	sick
11	57.0	male	asympt	140.0	192.0	f	normal	148.0	no	0.4	flat	0.0	fixed_defect	healthy
12	56.0	female	atyp_angina	140.0	294.0	f	left_vent_hyper	153.0	no	1.3	flat	0.0	normal	healthy
13	56.0	male	non_anginal	130.0	256.0	t	left_vent_hyper	142.0	yes	0.6	flat	1.0	fixed_defect	sick
14	44.0	male	atyp_angina	120.0	263.0	f	normal	173.0	no	0.0	up	0.0	reversable_defect	healthy
15	52.0	male	non_anginal	172.0	199.0	t	normal	162.0	no	0.5	up	0.0	reversable_defect	healthy
16	57.0	male	non_anginal	150.0	168.0	f	normal	174.0	no	1.6	up	0.0	normal	healthy
17	48.0	male	atyp_angina	110.0	229.0	f	normal	168.0	no	1.0	down	0.0	reversable_defect	sick
18	54.0	male	asympt	140.0	239.0	f	normal	160.0	no	1.2	up	0.0	normal	healthy
19	48.0	female	non_anginal	130.0	275.0	f	normal	139.0	no	0.2	up	0.0	normal	healthy
20	49.0	male	atyp_angina	130.0	266.0	f	normal	171.0	no	0.6	up	0.0	normal	healthy

Selain itu juga alasan peneliti menggunakan Data Cleaning secara manual yang mempengaruhi tingkat akurasinya, dan bisa dilihat hasil eksperimen pada tabel 3.4

Tabel 3.4 Perbandingan akurasi data sebelum cleaning dan data sesudah cleaning

	Accuracy	**AUC**
C4.5 (data = 867)	84,44	0,914
C4.5 (data = 567)	86,59	0,957

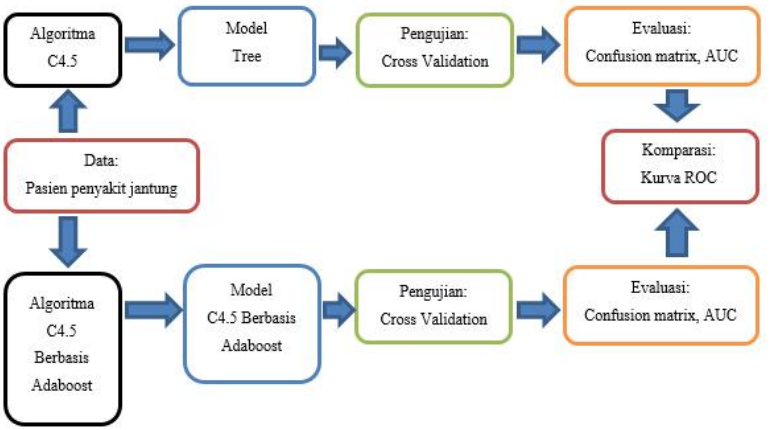
Dan dibawah ini adalah data yang memiliki atribut kosong, yang diabaikan dalam penelitian ini, bisa dilihat pada table 3.5, yaitu sebagai berikut:

Tabel 3.5 Data yang memiliki atribut kosong

no	umur	jenis kelamin	jenis sakitdada	tekanan darah	kolestrol	kadar gula	elektrokardiografi	tekanan jantung	angina induksi	oldpeak	segmen_st	flaurosopy	denyut jantung	hasil
1	53.0	female	non_anginal	128.0	216.0	f	left_vent_hyper	115.0	no	0.0	up	0.0	?	healthy
2	52.0	male	non_anginal	138.0	223.0	f	normal	169.0	no	0.0	up	?	normal	healthy
3	43.0	male	asympt	132.0	247.0	t	left_vent_hyper	143.0	yes	0.1	flat	?	reversable_defect	sick
4	52.0	male	asympt	128.0	204.0	t	normal	156.0	yes	1.0	flat	0.0	?	sick
5	58.0	male	atyp_angina	125.0	220.0	f	normal	144.0	no	0.4	flat	?	reversable_defect	healthy
6	38.0	male	non_anginal	138.0	175.0	f	normal	173.0	no	0.0	up	?	normal	healthy
7	38.0	male	non_anginal	138.0	175.0	f	normal	173.0	no	0.0	up	?	normal	healthy
8	28.0	male	atyp_angina	130.0	132.0	f	left_vent_hyper	185.0	no	0.0	?	?	?	healthy
9	29.0	male	atyp_angina	120.0	243.0	f	normal	160.0	no	0.0	?	?	?	healthy
10	29.0	male	atyp_angina	140.0	?	f	normal	170.0	no	0.0	?	?	?	healthy
11	30.0	female	atyp_angina	170.0	237.0	f	st_t_wave_abnormality	170.0	no	0.0	?	?	fixed_defect	healthy
12	31.0	female	atyp_angina	100.0	219.0	f	st_t_wave_abnormality	150.0	no	0.0	?	?	?	healthy
13	32.0	female	atyp_angina	105.0	198.0	f	normal	165.0	no	0.0	?	?	?	healthy
14	32.0	male	atyp_angina	110.0	225.0	f	normal	184.0	no	0.0	?	?	?	healthy
15	32.0	male	atyp_angina	125.0	254.0	f	normal	155.0	no	0.0	?	?	?	healthy
16	33.0	male	non_anginal	120.0	298.0	f	normal	185.0	no	0.0	?	?	?	healthy
17	34.0	female	atyp_angina	130.0	161.0	f	normal	190.0	no	0.0	?	?	?	healthy
18	34.0	male	atyp_angina	150.0	214.0	f	st_t_wave_abnormality	168.0	no	0.0	?	?	?	healthy
19	34.0	male	atyp_angina	98.0	220.0	f	normal	150.0	no	0.0	?	?	?	healthy
20	35.0	female	atyp_angina	120.0	160.0	f	st_t_wave_abnormality	185.0	no	0.0	?	?	?	healthy

3.4 Metode yang diusulkan

Model yang diusulkan pada penelitian ini adalah menggunakan algoritma *C4.5* dan algoritma *C4.5* berbasis *adaboost* yaitu:



Gambar 3.2 Model yang diusulkan

Pada gambar 3.2. menunjukan proses yang dilakukan dalam tahap modeling untuk menyelesaikan prediksi penyakit jantung dengan menggunakan dua metode yaitu algoritma *C4.5* dan algoritma *C4.5* berbasis *adaboost*.

1. Algoritma *C4.5* digunakan untuk membangun sebuah pohon keputusan yang mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar.
2. *Adaboost yaitu* metode untuk meningkatkan ketelitian dalam proses klasifikasi dan prediksi dengan cara membangkitkan kombinasi dari sebuah model, tetapi hasil klasifikasi atau prediksi yang di pilih adalah metode yang memiliki nilai bobot besar.

BAB IV
PEMBAHASAN

4.1 Eksperimen dan Pengujian Model

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan rule-rule yang ada pada model algoritma yang diusulkan. Setelah itu, diuji rule tadi kedalam model *cross validation* untuk mendapatkah hasil yang lebih baik.

4.1.1 Model Algoritma C4.5

Algoritma C4.5 untuk model yang pertama dilakukan. Berikut langkah-langkah yang akan dilakukan sebagai berikut:

1. Menghitung jumlah kasus class SICK dan class HEALTHY serta nilai *Entropy* dari semua kasus. Kasus dibagi berdasarkan atribut dengan jumlah kasus 567 *record*, kelas SICK ada 257 *record* dan kelas HEALTHY sebanyak 310 *record* sehingga didapat *entropy*.

= (-257/567. log2 (257/567)) + (-310/567. log2 (310/567))

= 0.9936

2. Hitung nilai Gain dari masing-masing atribut sebagai contoh untuk kolestrol:
   ≤ 311 = 516/567
   >311 = 51/567
   Atribut kolestrol ≤ 311 terdiri dari 231 class SICK dan 285 class HEALTHY, Atribut kolestrol > 311 terdiri dari 26 class

SICK dan 25 class HEALTHY Nilai Entropynya dapat dihitung sebagai berikut:

Kolestrol ≤ 311 = ((-231/516. log2 (231/516) +
(-285/516. log2 (285/516))
= 0.9920

Kolestrol > 311 = ((-26/51. log2 (26/51) +
(-25/51. log2 (25/51))
= 0.9997

E split kolestrol = ((516/567 (0.9920) +
(51/567 (0.9997))
= 0.9927

Gain kolestrol = 0.9936 – 0,9927
= 0.0009

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai gain tertinggi. Hasil perhitungan seluruh atribut adalah sebagai berikut:

Tabel 4.1 Informasi Gain untuk Algoritma C4.5

Candidate split	Kasus	SICK	HEALTHY	Entropy	Gian
Denyut_jantung					
= fixed_defect	32	20	12	0,0539	
= normal	315	69	246	0,4213	
= reversable_defect	220	168	52	0,3061	
Flaurosofy	334	85	249	0,4820	
= 0	123	82	41	0,1992	
= 1	71	57	14	0,0897	
= 2	39	33	6	0,0426	
= 3	149	85	64	0,9795	0,0142
	418	172	246		

Kolestrol > 274 ≤ 274 >182 ≤ 182	525 42	238 19	287 23	0,9937	0

Selengkapnya terdapat pada lampiran 1

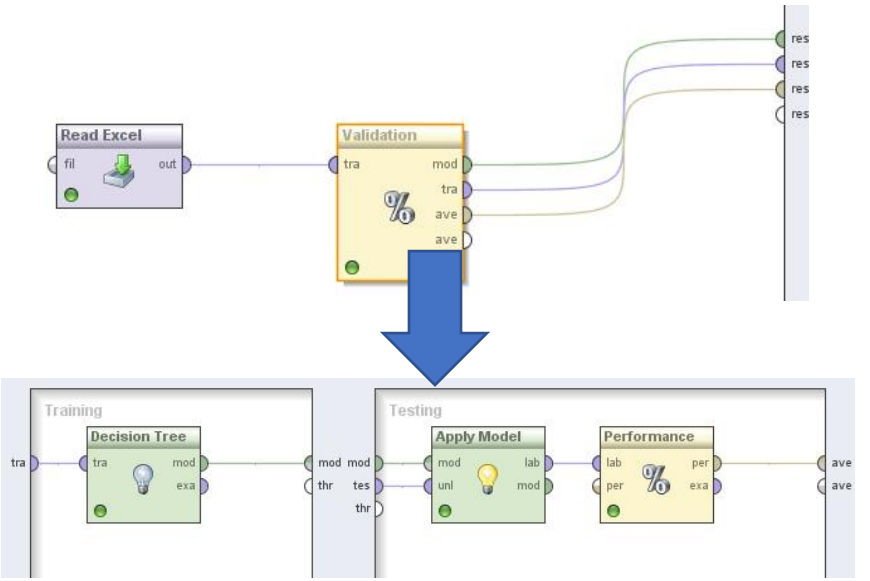
Dari hasil perhitungan diatas, maka dapat digambarkan model pohon keputusan Algoritma C4.5 yang terdapat pada lampiran 2.

Tujuan utama dari menganalisis data dengan menggunakan algoritma *C4.5* ini adalah ingin mendapatkan rule (Witten 2007), yang akan dimanfaatkan untuk pengambilan keputusan pada data baru. Adapun rule yang diapat dari Gambar 4.1 dan gambar 4.2 diatas adalah:

1. R1: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND kolestrol >271 THEN class = SICK
2. R2: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND kolestrol ≤271 THEN class = HEALTHY
3. R3: IF denyut_jantung = fixed_defect AND flaurosopy = 1 THEN class = SICK
4. R4: IF denyut_jantung = fixed_defect AND flaurosopy = 2 THEN class = SICK
5. R5: IF denyut_jantung = fixed_defect AND flaurosopy = 3 THEN class = SICK

   Dan selengkapnya terdapat pada lampiran 3.

Berikut adalah gambar pengujian menggunakan motode K-*Fold Cross Validation*:



Gambar 4.1 Pengujian K-*Fold Cross Validation* Algoritma C4.5

Dalam pengujian K-*Fold Cross Validation* Algoritma C4.5, peneliti menggunakan 10 kali percobaan dengan sampling type Stratified (bertingkat-tingkat) dengan menggunakan use local random seed karena hasil akurasinya lebih tinggi. Dan dibawah ini tabel perbandingannya.

Tabel 4.2 Perbandingan sampling type Stratified

Cross Validation	Random	Non random
Akurasi	86,59%	85,47%

4.1.2 Algoritma C4.5 berbasis Adaboost

Algoritma C4.5 berbasis Adaboost untuk model yang kedua membuat iterasi sebanyak seuluh kali dan menghasilkan 10 arsitektur algoritma *C4.* p5 dengan nilai w berbeda yaitu:

1. Model 1 [w = 4.538]
2. Model 2 [w = 5.409]

3. Model 3 [w = 5.309]
4. Model 4 [w = 2.979]
5. Model 5 [w = 5.965]
6. Model 6 [w = 1.103]
7. Model 7 [w = 3.701]
8. Model 8 [w = 1.672]
9. Model 9 [w = 1.748]
10. Model 10 [w = 9.586]

Tiap model tersebut akan melalui langkah-langkah yang akan dilakukan sebagai berikut:

1. Menghitung jumlah kasus class SICK dan class HEALTHY serta nilai *Entropy* dari semua kasus. Kasus dibagi berdasarkan atribut dengan jumlah kasus 567 *record*, kelas SICK ada 257 *record* dan kelas HEALTHY sebanyak 310 *record* sehingga didapat *entropy*.

= (-257/567. log2 (257/567))
+ (-310/567. log2 (310/567))
= 0.9936

2. Hitung nilai Gain dari masing-masing atribut sebagai contoh untuk kolestrol:
   ≤ 301 = 483/567
   > 301 = 84/567
   Atribut kolestrol ≤ 301 terdiri dari 218 class SICK dan 265 class HEALTHY, Atribut kolestrol > 301 terdiri dari 36 class SICK dan 45 class HEALTHY Nilai Entropynya dapat dihitung sebagai berikut:

   Kolestrol ≤ 301 = ((-218/483. log2 (218/483)
   + (-265/483. log2 (265/483))
   = 0.9931

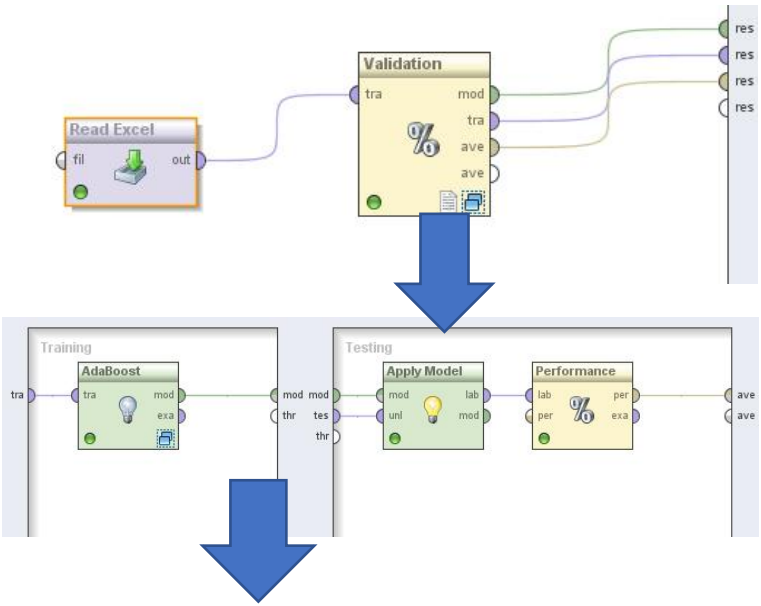
Kolestrol > 301 = ((-36/84. log2 (36/84)
+ (-45/84. log2 (45/84))
= 0.9963

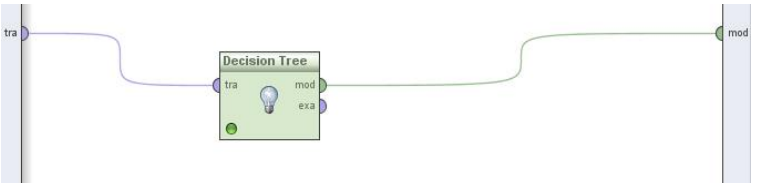
E split kolestrol = ((483/567 (0.9931)
+ (48/567 (0.9963))
= 0.9936

Gain kolestrol = 0.9936 – 0,9935
= 0.0001

Perhitungan *entropy* dan *gain* untuk semua atribut dilakukan, untuk mendapatkan nilai gain tertinggi. Hasil perhitungan seluruh atribut, gambar model pohon keputusan dan rule dari 10 model terdapat pada lampiran 4.

Berikut adalah gambar pengujian menggunakan motode K-*Fold Cross Validation*:





Gambar 4.2 Pengujian K-*Fold Cross Validation* Algoritma C4.5 berbasis Adaboost

Dalam pengujian K-*Fold Cross Validation* Algoritma C4.5 berbasis Adaboost, peneliti juga menggunakan 10 kali percobaan dengan sampling type Stratified (bertingkat-tingkat) dengan menggunakan use local random seed karena hasil akurasinya juga lebih tinggi. Dan dibawah ini tabel perbandingannya.

Tabel 4.3 Perbandingan sampling type Stratified

Cross Validation	Random	Non random
Akurasi	92,24%	91,84%

4.2 Evaluasi dan validasi Hasil

Metode klasifikasi bisa dievaluasi berdasarkan beberapa kriteria seperti tingkat akurasi, kecepatan, kehandalan, skalabilitas, dan interpretabilitas (Vercellis 2009). Hasil pengujian model yang dilakukan dalam bab tiga adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*) dari prediksi penyakit jantung dengan metode *cross validation*

4.2.1 Hasil Pengujian Model Algoritma

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*).

a. *Confusion Matrix*

Tabel dibawah ini adalah hasil pengujian dengan jumlah data 567 record. Berikut Tabel yang didapat:

Tabel 4.4 Model *Confusion Matrix* untuk Algoritma C4.5

accuracy: 86.59% +/- 4.12% (mikro: 86.60%)

	true healthy	true sick	class precision
pred. healthy	270	36	88.24%
pred. sick	40	221	84.67%
class recall	87.10%	85.99%	

Jumlah *True Positive* (TP) adalah 270 *record* diklasifikasikan sebagai HEALTHY terpilih dan *False Negative* (FN) sebanyak 38 *record* diklasifikasikan sebagai HEALTHY terpilih tetapi SICK terpilih. Berikutnya 221 *record* untuk *True Negative* (TN) diklasifikasikan sebagai SICK terpilih, dan 40 *record False Positive* (FP) diklasifikasin sebagai SICK terpilih ternyata HEALTHY. Berdasarkan Tabel 4.4 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 adalah sebesar 86,59%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \quad acc = \frac{270 + 221}{270 + 221 + 40 + 36}$$

$$Sensitivity = \frac{tp}{tp + fn} \quad Sensitivity = \frac{270}{270 + 36}$$

$$Specitivity = \frac{tn}{tn + fp} \quad Specitivity = \frac{221}{221 + 40}$$

$$ppv = \frac{tp}{tp + fp} \quad ppv = \frac{270}{270 + 40}$$

$$npv = \frac{tn}{tn + fn} \quad npv = \frac{221}{221 + 36}$$

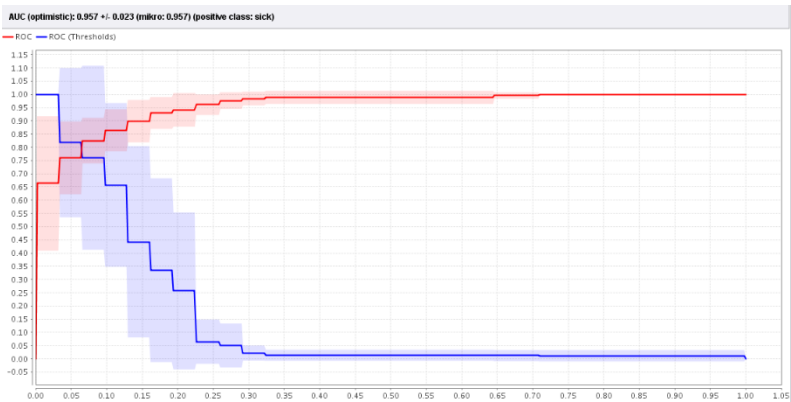
Hasil perhitungan dari persamaan diatas terlihat pada Tabel 4.5 dibawah ini:

Tabel 4.5 Nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv*

	Nilai (%)
Accuracy	86,59
Sensitivity	88,23
Spesificity	84,67
PPV	87,09
NPV	86

b. Evaluasi ROC *curve*

Dari Tabel 4.5 terdapat grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.957 dengan nilai akurasi *Excellent Classification*.



Gambar 4.3 Nilai AUC dalam grafik ROC algoritma C4.5

4.2.2 Hasil Pengujian Model Algoritma C4.5 berbasis Adaboost

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Curve*).

a. Counfusion Matrix

Tabel dibawah ini adalah hasil pengujian dengan jumlah data 567 *record*. Berikut Tabel yang didapat:

Tabel 4.6 Model *counfusion matrix* untuk Algoritma C4.5 berbasis Adaboost

accuracy: 92.24% +/- 4.17% (mikro: 92.24%)

	true healthy	true sick	class precision
pred. healthy	291	25	92.09%
pred. sick	19	232	92.43%
class recall	93.87%	90.27%	

Jumlah *True Positive* (TP) adalah 291 *record* diklasifikasikan sebagai HEALTHY terpilih dan *False Negative* (FN) sebanyak 25 *record* diklasifikasikan sebagai HEALTHY terpilih tetapi SICK terpilih. Berikutnya 232 *record* untuk *True Negative* (TN) diklasifikasikan sebagai SICK terpilih, dan 19 *record False Positive* (FP) diklasifikasin sebagai SICK terpilih ternyata HEALTHY. Berdasarkan Tabel 4.6 menunjukan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 berbasis *Adaboost* adalah sebesar 92,24%, dan dapat dihitung untuk mencari nilai *accuracy, sensitivity, specificity, ppv,* dan *npv* pada persamaan dibawah ini:

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \; acc = \frac{291 + 232}{291 + 232 + 19 + 25}$$

$$Sensitivity = \frac{tp}{tp + fn} \; Sensitivity = \frac{291}{291 + 25}$$

$$Specitivity = \frac{tn}{tn + fp} \; Specitivity = \frac{231}{231 + 19}$$

$$ppv = \frac{tp}{tp + fp} \; ppv = \frac{291}{291 + 19}$$

$$npv = \frac{tn}{tn + fn} \; npv = \frac{232}{232 + 25}$$

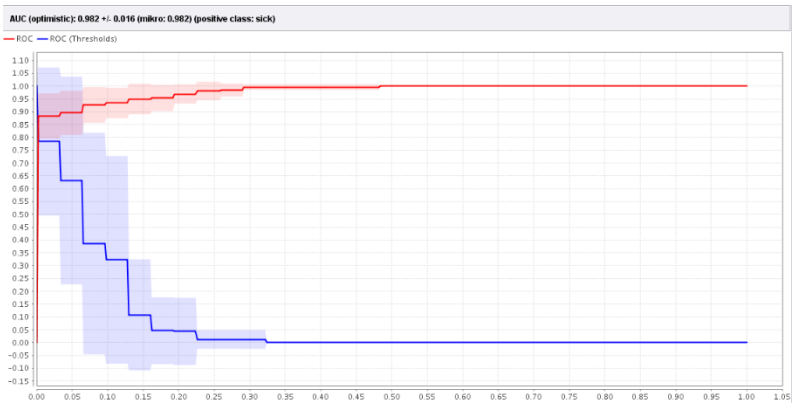
Hasil perhitungan dari persamaan diatas terlihat pada Tabel 4.7 dibawah ini:

Tabel 4.7 Nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv*

	Nilai (%)
Accuracy	92,24
Sensitivity	91,08
Spesificity	92,4
PPV	93,54
NPV	90,27

b. Evaluasi ROC *curve*

Gambar 4.4 menunjukan grafik ROC dengan nilai AUC (*Area Under Curve*) sebesar 0.982 dengan tingkat diagnosa *Excellent classification*.



Gambar 4.4 Nilai AUC dalam grafik ROC algoritma C4.5 berbasis Adaboost

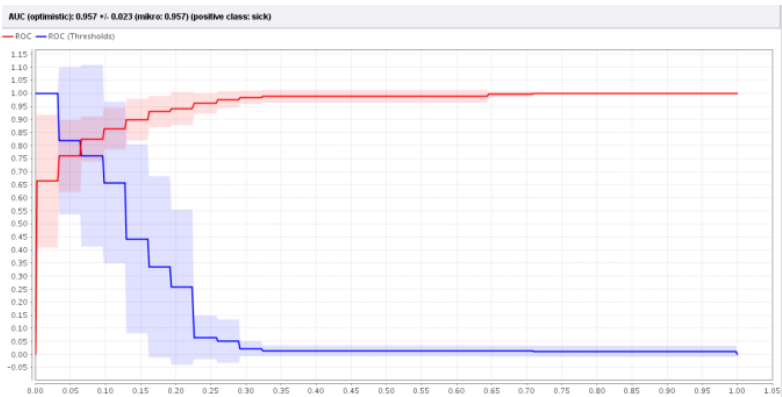
4.2.3 Analisis Evaluasi dan Validasi Model

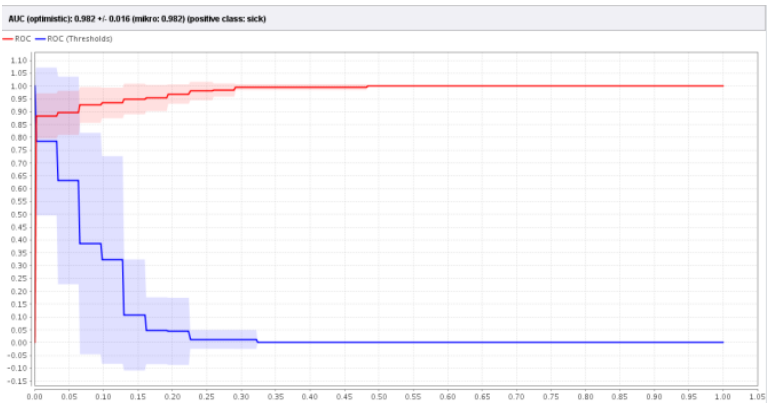
Dari hasil pengujian diatas, baik evaluasi menggunakan *counfusion matrix* maupun ROC *curve* terbukti bahwa hasil pengujian algoritma C4.5 berbasis Adaboost memiliki nilai akurasi yang lebih tinggi dibandingkan dengan algoritma C4.5. Nilai akurasi untuk model algoritma C4.5 sebesar 86,59% dan nilai akurasi untuk model algoritma C4.5 berbasis Adaboost sebesar 92.24% dengan selisih akurasi 5,65%, dapat dilihat pada Tabel 4.8 dibawah ini:

Tabel 4.8 Pengujian algoritma C4.5 dan C4.5 berbasis Adaboost

	Accuracy	**AUC**
C4.5	86,59	0,957
C4.5 Berbasis Adaboost	92,24	0,982

Untuk evaluasi menggunakan ROC *curve* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma C4.5 menghasilkan nilai 0.957 dengan nilai diagnosa *Excellent Classification*, sedangkan untuk algoritma C4.5 berbasis Adaboost menghasilkan nilai 0.982 dengan nilai diagnose *Excellent Classification*, dan selisih nilai keduanya sebesar 0.025. Dapat dilihat pada Gambar 4.5 dibawah ini.





Gambar 4.5 ROC *curve* (Algoritma C4.5 dan Algoritma C4.5 berbasis Adaboost)

Melihat tabel 4.8 dan gambar 4.5, maka dapat dihitung peningkatan akurasi dari pengembangan Algoritma C4.5 ke Algoritma C4.5 berbasis Adaboost dengan perhitungan dibawah ini:

$$Tingkat\ akurasi\ (\%) = \frac{SS - SB}{SB} x100$$

$$Tingkat\ akurasi\ (\%) = \frac{5{,}65}{86{,}59} x100$$

$$Tingkat\ akurasi\ (\%) = 6{,}42\%$$

Jadi akurasi dari Algoritma C4.5 ke Algoritma C4.5 berbasis Adaboost meningkat 6,42%.

Sedangkan untuk peningkatan kurva ROC dari Algoritma C4.5 ke Algoritma C4.5 berbasis Adaboost dengan perhitungan dibawah ini:

$$Tingkat\ AUC\ (\%) = \frac{SS - SB}{SB} x100$$

$$Tingkat\ AUC\ (\%) = \frac{0.025}{0{,}957} x100$$

$$Tingkat\ AUC\ (\%) = 0{,}26\%$$

Jadi nilai AUC dari Algoritma C4.5 ke Algoritma C4.5 berbasis Adaboost meningkat 0,26%.

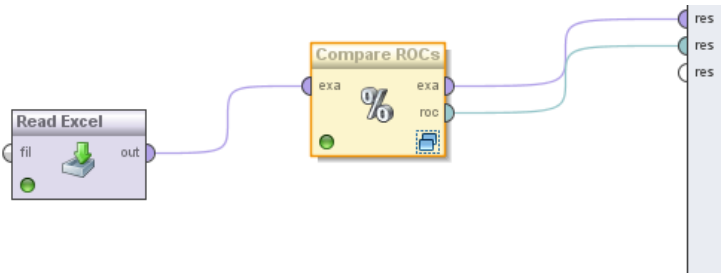
Selain itu juga peneliti membandingkan Algoritma C4.5 berbasis Adaboost dengan Algoritma C4.5 berbasis algoritma optimasi lain seperti Bagging. Hasil pengujiannya adalah sebagai:

Tabel 4.9 Pengujian algoritma C4.5 dan C4.5 berbasis Adaboost & Bagging

	Accuracy	**AUC**
C4.5	86,59	0,957
C4.5 Berbasis Adaboost	92,24	0,982
C4.5 Berbasis Bagging	91,89	0,962

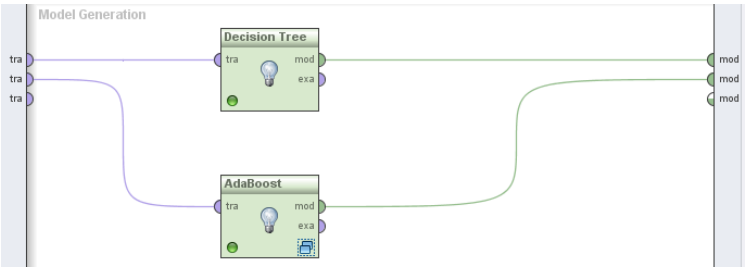
Melihat dari hasil pengujian diatas, Algoritma *C4.5* berbasis *Adaboost* menghasilkan akurasi tertinggi yaitu 92,24% dibandingkan dengan Algoritma *C4.5* berbasis *Bagging* dengan akurasi 91,89%, dengan selisih 0,35%.

Selain pengujian *performace* dengan menggunakan *confusion matrix*. Data penyakit jantung akan dilakukan pengujian menggunakan komparasi dengan dengan menggunakan *ROC curve*. Berikut desain dengan menggunakan desain model evaluasi komparasi dengan mengunakan *ROC Curve* dengan menggunakan framework RapidMiner versi 5.2.008.



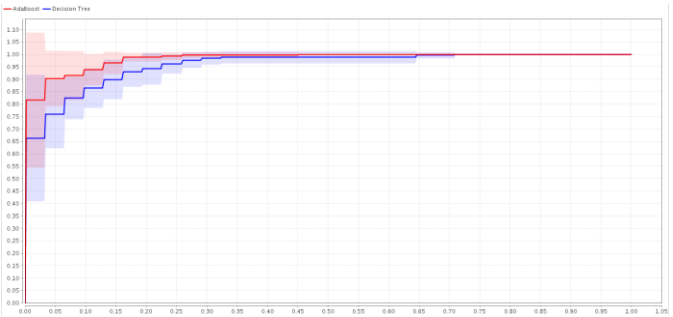
Gambar 4.6 Desain model komparasi menggunakan *ROC Curve*

Sedangkan untuk model *ROC curve* seperti berikut:



Gambar 4.7 Model komparasi *ROC Curve*

Maka akan dihasilkan grafik hasil eksekusi model *ROC curve* seperti dibawah ini:



Gambar 4.8 Komparasi *ROC Curve* pada algoritma *C4.5* dan algoritma *C4,5* berbasis *adaboost*

Dari hasil perhitungan model pada table 4.8 dengan penerapan klasifikasi *performance* keakurasian AUC maka dapat diklasifikasikan menjadi lima kelompok (Gorunescu 2011), antara lain:

a. 0.50 – 0.60 = klasifikasi salah

b. 0.60 – 0.70 = klasifikasi buruk

c. 0.70 – 0.80 = klasifikasi cukup

d. 0.80 – 0.90 = klasifikasi baik

e. 0.90 – 1.00 = klasifikasi sangat baik

Jadi berdasarkan pengelompokan diatas pada table 4.1 dengan membandingkan nilai *accuracy* dan AUC terlihat bahwa algoritma *C4.5* berbasis *adaboost* memiliki nilai accuracy dan nilai AUC yang lebih baik dibandingkan algoritma *C4.5* dan dapat disimpulkan bahwa nilai AUC *C4.5* dan *C4.5* berbasis *adaboost* antara 0.90 – 1.00 termasuk klasifikasi sangat baik.

BAB V
PENUTUP

5.1 Kesimpulan

Dalam penelitian ini dilakukan pengujian model dengan menggunakan algoritma *C4.5* dan algoritma *C4.5* berbasis *adaboost* dengan menggunakan data pasien yang menderita penyakit jantung atau tidak. Model yang dihasilkan diuji untuk mendapatkan nilai *accuracy*, dan AUC dari setiap algoritma sehingga didapat pengujian dengan menggunakan *C4.5* didapat nilai *accuracy* adalah 86,59 % dengan nilai AUC adalah 0.957. sedangkan pengujian dengan mengunakan *C4.5* berbasis *adaboost* didapatkan nilai *accuracy* 92.24 % dengan nilai AUC adalah 0.982. selain itu juga peneliti mengkomparasi dengan algoritma C4.5 berbasis Bagging didapat *accuracy* 91,89% dan nilai AUC 0,963. Maka dapat disimpulkan pengujian model penyakit jantung dengan menggunakan algoritma *C4.5* berbasis *adaboost* lebih baik dari pada C4.5 sendiri, dengan peningkatan akurasi sebesar 6,42% dan peningkatan nilai AUC sebesar 0,26%.

Dengan demikian dari hasil pengujian model diatas dapat disimpulkan bahwa *C4.5* berbasis *adaboost* memberikan pemecahan untuk permasalahan penyakit jantung lebih akurat.

5.2 Saran

Dari hasil pengujian yang telah dilakukan dan hasil kesimpulan yang diberikan maka ada saran atau usul yang di berikan antara lain:

1. Dalam Penelitian ini dilakukan menggunakan metode algoritma *C4.5* dan algoritma C4.5 berbasis metode *Adaboost*. Mencoba mengurangi beberapa atribut dan mencobakan kembali dengan algoritma lain dengan mengoptimalisasi selain ada boost yang menghasilkan tingkat akurasi tinggi.
2. Hasil penelitian ini diharapkan bisa digunakan untuk rumah sakit untuk meningkatkan akurasi dalam prediksi penyakit jantung.

DAFTAR PUSTAKA

Annisa, Riski. 2019. "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung." *Jurnal Teknik Informatika Kaputama (JTIK)* 3 (1): 22–28. https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156.

Aulia, Wizra. 2018. "Sistem Pakar Diagnosa Penyakit Jantung Koroner Dengan Metode Probabilistic Fuzzy Decision Tree." *Jurnal Sains Dan Informatika* 4 (2): 106. https://doi.org/10.22216/jsi.v4i2.3258.

Bisri, Achmad, and Romi Satria Wahono. 2015. "Penerapan Adaboost Untuk Penyelesaian Ketidakseimbangan Kelas Pada Penentuan Kelulusan Mahasiswa Dengan Metode Decision Tree." *Journal of Intelligent Systems* 1 (1): 27–32.

Budi Santosa. 2007. *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.

Derisma, D. 2020. "Perbandingan Kinerja Algoritma Untuk Prediksi Penyakit Jantung Dengan Teknik Data Mining." *Journal of Applied Informatics and Computing* 4 (1): 84–88. https://doi.org/10.30871/jaic.v4i1.2152.

Fadli, Ari. 2011. "Konsep Data Mining," 1–9. www.ilmukomputer.com.

Gorunescu, Florin. 2011. *Data Mining: Concepts, Models and Techniques (Intelligent Systems Reference Library)*. *Soft Computing*.

Iskandar, Iskandar, Abdul Hadi, and Alfridsyah Alfridsyah. 2017. "Faktor Risiko Terjadinya Penyakit Jantung Koroner Pada Pasien Rumah Sakit Umum Meuraxa Banda Aceh." *AcTion: Aceh Nutrition Journal* 2 (1): 32. https://doi.org/10.30867/action.v2i1.34.

K, Kasikumar, Mohamed Najumuddeen M, and Suresh R. 2018. "Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks." *International Journal of Data Mining Techniques and Applications* 7 (1): 172–76. https://doi.org/10.20894/ijdmta.102.007.001.027.

Koroner, Jantung, Pada Usia, Dewasa Madya, Studi Kasus, Umum Daerah, Kota Semarang, Amelia Farahdika, and Mahalul Azam. 2015. "FAKTOR RISIKO YANG BERHUBUNGAN DENGAN PENYAKIT JANTUNG KORONER PADA USIA DEWASA MADYA (41-60 TAHUN) (Studi Kasus Di RS Umum Daerah Kota Semarang)." *Unnes Journal of Public Health* 4 (2): 117–23. https://doi.org/10.15294/ujph.v4i2.5188.

M Anbarasi, E Anupriya, N.Ch.S.N. Iyengar. 2010. "Enhanced Prediction of Heart Disease with Feature Subset Selection Using Genetic Algorithm Enhanced Prediction of Heart Disease with Feature Subset Selection Using Genetic Algorithm." *International Journal of Engineering Science and Technology* 2 (10): 5370–76.

Olusola, Adetunmbi A, Adeola S Oladele, and Daramola O Abosede. 2016. "Analysis of KDD & Apos ; 99 Intrusion

Detection Dataset for Selection of Relevance Features Analysis of KDD ' 99 Intrusion Detection Dataset for Selection of Relevance Features" I (January): 16–23.

Parlar, Tuba, and Songul Kakilli Acaravci. 2017. "International Journal of Economics and Financial Issues Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data." *International Journal of Economics and Financial Issues* 7 (2): 692–96. http:www.econjournals.com.

Pratiwi, Shiela Novelia Dharma, and Brodjol Sutijo Suprih Ulama. 2016. "Klasifikasi Email Spam Dengan Menggunakan Metode Support Vector Machine Dan K-Nearest Neighbor." *Jurnal Sains Dan Seni ITS* 5 (2): 344–49. https://doi.org/10.12962/j23373520.v5i2.16685.

Rohman, Abdul. 2017. "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Jantung." *Neo Teknika* 2 (2): 21–28. https://doi.org/10.37760/neoteknika.v2i2.766.

Salvi, Vinita. 2016. "Heart Disease." *Handbook of Obstetrics and Gynecology for Asia and Oceania*, 272–272. https://doi.org/10.5005/jp/books/12747_43.

Soysal, Murat, and Ece Guran Schmidt. 2010. "Machine Learning Algorithms for Accurate Flow-Based Network Traffic Classification: Evaluation and Comparison." *Performance Evaluation* 67 (6): 451–67. https://doi.org/10.1016/j.peva.2010.01.001.

Suwondo, Adi, Dian Asmarajati, and Heri Surahman. 2013. "Algoritma C4.5 Berbasis Adaboost Untukprediksi Penyakit Jantung Koroner." *Seminar Nasional Teknologi Dan Teknopreneur (SNTT)*, 1–11. http://id.scribd.com/doc/178672740/Prossiding.

Utomo, Dito Putro, Pahala Sirait, and Roni Yunis. 2020. "Reduksi Atribut Pada Dataset Penyakit Jantung Dan Klasifikasi Menggunakan Algoritma C5. 0." *Jurnal Media Informatika Budidarma* 4 (4): 994–1006. https://doi.org/10.30865/mib.v4i4.2355.

Vercellis, Carlo. 2009. *Business Intelligence: Data Mining and Optimization for Decision Making*. *Business Intelligence: Data Mining and Optimization for Decision Making*. https://doi.org/10.1002/9780470753866.

Vulandari, Retno Tri. 2017. *Data Mining Teori Dan Aplikasi Rapidminer*. Yogyakarta: Gava Media.

Witten, Ian H. 2007. "Data Mining Data Mining Complications : Overfitting Statistical Modeling One Attribute Does All the Work ?"

GLOSARIUM

Algoritma : prosedur sistematis untuk memecahkan masalah matematis dalam langkah-langkah terbatas

Adaboost : metode ansambel berulang. Itu membangun pengklasifikasi yang kuat dengan menggabungkan beberapa pengklasifikasi yang berkinerja lemah.

Akurasi : ukuran kedekatan hasil pengukuran dengan nilai sebenarnya atau nilai target.

Angina : jenis sakit dada

C4.5 :algoritma yang digunakan untuk menghasilkan pohon keputusan yang dikembangkan oleh Ross Quinlan

Confusion Matrix : tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual.

Cross Validation : sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independen

Data Cleaning : proses analisa mengenai kualitas dari data dengan mengubahnya

Decision Tree :alat pendukung keputusan yang menggunakan model keputusan seperti pohon dan kemungkinan

konsekuensinya, termasuk hasil acara kebetulan, biaya sumber daya, dan utilitas.

Data Training : Untuk melatih algoritma

Data Set :dipakai untuk mengetahui performa algoritma yang sudah dilatih

Data Mining :proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar.

Eksperimen :Cara untuk mencari hubungan sebab akibat (hubungan kausal) anatar dua faktor yang sengaja ditimbulkan oleh peneliti

Kurva ROC : cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual

Stratified Type: Suatu teknik pengambilan sampel dengan memperhatikan suatu tingkatan (strata) pada elemen populasi.

INDEKS

LAMPIRAN-LAMPIRAN

Lampiran 1. Tabel Informasi Gain untuk Algoritma C4.5

Candidate split	Kasus	SICK	HEALTHY	Entropy	Gian
Denyut_jantung					
fixed_defect	32	20	12	0,0539	
normal	315	69	246	0,4213	
reversable_defect	220	168	52	0,3061	
Flaurosofy	334	85	249	0,4820	
= 0	123	82	41	0,1992	
= 1	71	57	14	0,0897	
= 2	39	33	6	0,0426	
= 3	149	85	64	0,9795	0,0142
Kolestrol	418	172	246		
> 274	525	238	287	0,9937	0
≤ 274	42	19	23		
>182	51	26	25	0,9928	0,0009
≤ 182	516	231	285		
> 311	410	203	207	0,9802	0,0135
≤ 311	157	54	103		
> 215	45	24	21	0,9921	0,0016
≤ 215	522	233	289		
> 316	10	4	6	0,9935	0,0001

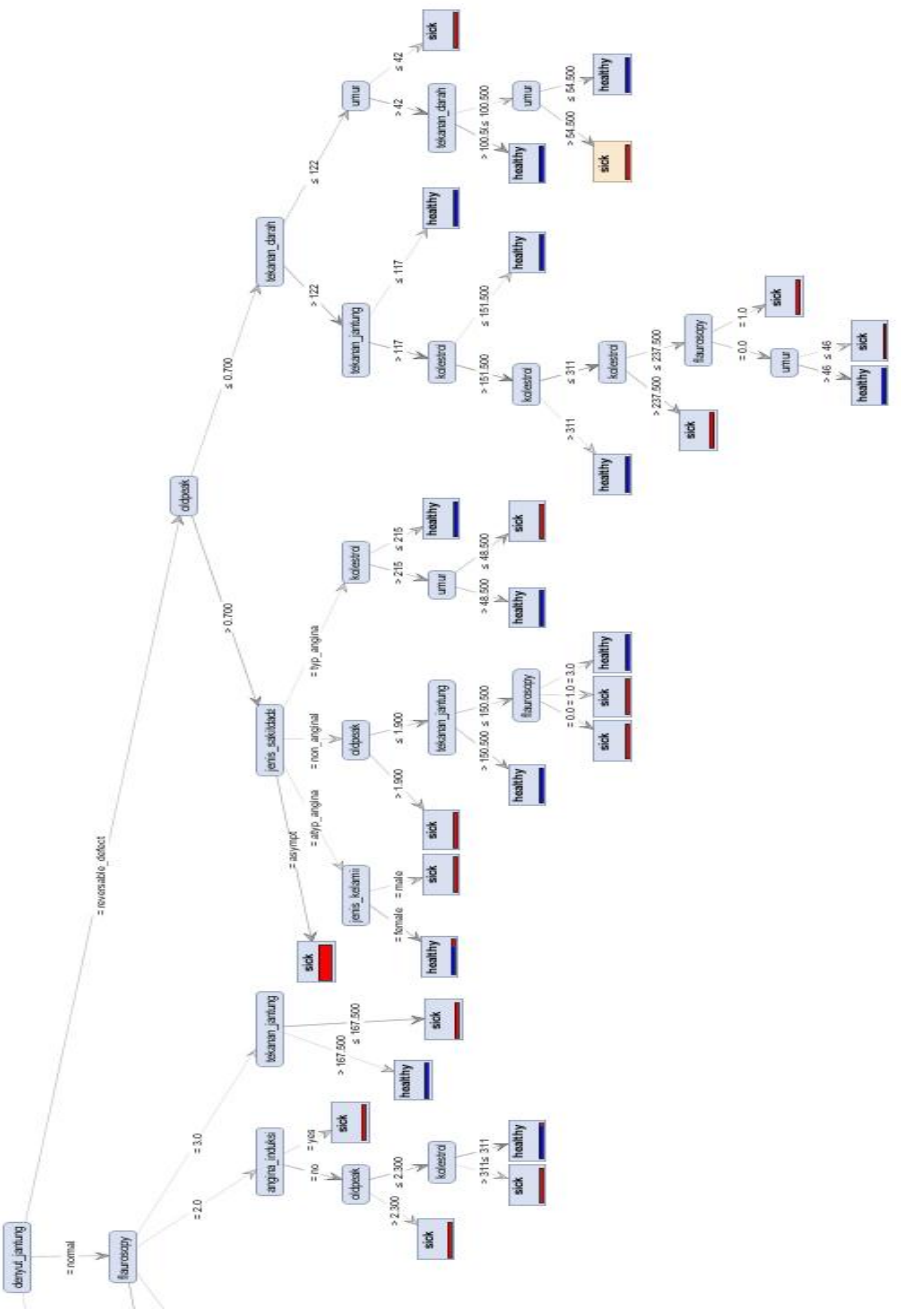
≤ 316	557	253	304		
> 364,500	283	148	135	0,9795	0,0141
≤ 364,500	284	109	175		
> 243,500	558	254	304	0,9930	0,0007
≤ 243,500	9	3	6		
> 151,500	309	157	152	0,9832	0,0105
≤ 151,500	258	100	158		
> 237,500					
≤ 237,500	54	33	21	0,9861	0,0076
Tekanan_darah	513	224	289		
> 157	531	249	282	0,9824	0,0113
≤ 157	36	8	28		
> 109	543	253	290	0,9820	0,0117
≤ 109	24	4	20		
> 106,500	337	158	179	0,9927	0,0010
≤ 106,500	230	99	131		
> 126,500	199	67	52	0,9844	0,0093
≤ 126,500	448	190	258		
> 144,500	376	185	191	0,9850	0,0087
≤ 144,500	191	72	119		
> 122	559	253	306	0,9936	0,0001
≤ 122	8	4	4		
> 100,500					
≤ 100,500	114	56	58	0,9926	0,0011
Umur	453	210	252		
> 62	174	100	74	0,9747	0,0190
≤ 62	393	157	236		
> 59,500	296	174	122	0,9352	0,0585
≤ 59,500	271	83	188		
> 54,500	412	212	200	0,9638	0,0299
≤ 54,500	155	45	110		
> 48,500	436	223	213	0,9595	0,0342
≤ 48,500	131	34	97		
> 46	500	241	259	0,9747	0,0190
≤ 46	67	16	52		
> 42					

≤ 42	14	12	2	0,9808	0,0129
Tekanan_jantung	553	245	308		
> 83,500	239	69	170	0,9350	0,0587
≤ 83,500	328	188	140		
> 159,500	18	14	4	0,9833	0,0104
≤ 159,500	249	243	306		
> 199,500	90	25	65	0,9761	0,0176
≤ 199,500	477	232	245		
> 173	153	41	112	0,9554	0,0383
≤ 173	414	216	198		
> 167,500	324	97	227	0,9002	0,0935
≤ 167,500	243	160	83		
> 150,500	520	220	300	0,9633	0,0304
≤ 150,500	47	37	10		
> 117					
≤ 117	48	42	6	0,9418	0,0518
Oldpeak	519	215	304		
> 2,800	24	22	2	0,9627	0,0310
≤ 2,800	543	235	308		
> 3,500	141	113	28	0,8722	0,1215
≤ 3,500	426	144	282		
> 1,700	80	70	10	0,9019	0,0918
≤ 1,700	487	187	300		
> 2,300	292	187	105	0,8822	0,1115
≤ 2,300	275	70	205		
> 0,700	111	93	18	0,8831	0,1106
≤ 0,700	456	164	292		
> 1,900					
≤ 1,900	380	116	264	0,5949	
Angina_induksi	187	141	46	0,2654	
= no					
= yes	23	7	16	0,0360	
Jenis_sakit dada	71	15	56	0,0932	
= typ_angina	162	35	127	0,2151	
= atyp_angina	311	200	111	0,5156	
= non_anginal					

= asympt					

Sehingga pohon keputusan dari hasil perhitungan diatas dapat digambarkan pada Gambar dibawah ini:

Lampiran 2. Gambar Model Pohon Keputusan Algoritma C4.5



Lampiran 3. Rule Pohon Keputusan Algoritma C4.5

1. R1: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND kolestrol >271 THEN class = SICK
2. R2: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND kolestrol ≤271 THEN class = HEALTHY
3. R3: IF denyut_jantung = fixed_defect AND flaurosopy = 1 THEN class = SICK
4. R4: IF denyut_jantung = fixed_defect AND flaurosopy = 2 THEN class = SICK
5. R5: IF denyut_jantung = fixed_defect AND flaurosopy = 3 THEN class = SICK
6. R6: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah >157 AND umur > 62 THEN class = HEALTHY
7. R7: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah >157 AND umur ≤ 62 THEN class = SICK
8. R8: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak > 2.8 THEN class = SICK
9. R9: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak ≤ 2.8 AND kolestrol > 316 AND kolestrol > 364.5 THEN class = HEALTHY
10. R10: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak ≤ 2.8 AND kolestrol > 316 AND kolestrol ≤ 364.5 THEN class = SICK
11. R11: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak ≤ 2.8 AND kolestrol > 316 AND angina_induksi=no THEN class = HEALTHY

12. R12: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak ≤ 2.8 AND kolestrol > 316 AND angina_induksi=yes AND kolestrol >243.5 THEN class = SICK
13. R13: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung > 83.5 AND oldpeak ≤ 2.8 AND kolestrol > 316 AND angina_induksi=yes AND kolestrol ≤243.5 THEN class = HEALTHY
14. R14: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur > 59.5 AND tekanan_jantung ≤83.5 THEN class = SICK
15. R15: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak >3.55 THEN class = SICK
16. R16: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND umur > 54 THEN class = SICK
17. R17: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND umur ≤ 54 THEN class = HEALTHY
18. R18: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND tekanan_darah > 109 THEN class = HEALTHY
19. R19: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND tekanan_darah ≤ 109 AND tekanan_darah ≤ 106,5 AND tekanan_jantung > 159,5 THEN class = HEALTHY

20. R20: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND tekanan_darah ≤ 109 AND tekanan_darah ≤ 106,5 AND tekanan_jantung ≤ 159,5 THEN class = SICK
21. R21: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_darah ≤ 157 AND umur ≤ 59.5 AND oldpeak ≤ 3.55 AND oldpeak ≤ 1.7 AND tekanan_darah ≤ 109 AND tekanan_darah ≤ 106,5 THEN class = SICK
22. R22: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung > 173 THEN class = SICK
23. R23: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung ≤ 119,5 THEN class = SICK
24. R24: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol >182 AND jenis_sakit dada = asympt AND tekanan_darah >126,5 THEN class = HEALTHY
25. R25: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol >182 AND jenis_sakit dada = atyp_angina AND tekanan_darah ≤126,5 THEN class = SICK
26. R26: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol >182 AND jenis_sakit dada = atyp_angina AND tekanan_darah >144 THEN class = SICK
27. R27: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol >182 AND jenis_sakit dada = atyp_angina AND tekanan_darah ≤144 THEN class = SICK
28. R28: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND

kolestrol >182 AND jenis_sakit dada = non_anginal THEN class = HEALTHY

29. R29: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol >182 AND jenis_sakit dada = typ_angina THEN class = HEALTHY
30. R30: IF denyut_jantung = normal AND flaurosopy = 0 AND tekanan_jantung >119,5 AND tekanan_jantung ≤ 173 AND kolestrol ≤ 182 THEN class = HEALTHY
31. R31: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah ≤ 122 AND umur ≤ 42 THEN class = SICK
32. R32: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah ≤ 122 AND umur > 42 AND tekanan_darah > 100.5 THEN class = HEALTHY
33. R33: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah ≤ 122 AND umur > 42 AND tekanan_darah ≤ 100.5 umur > 54.5 THEN class = SICK
34. R34: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah ≤ 122 AND umur > 42 AND tekanan_darah ≤ 100.5 umur ≤ 54.5 THEN class = HEALTHY
35. R35: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117 AND kolestrol > 151.5 AND kolestrol >311 THEN class = HEALTHY
36. R36: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung ≤ 117 THEN class = HEALTHY
37. R37: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117 AND kolestrol ≤ 151.5 THEN class = HEALTHY
38. R38: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117

AND kolestrol > 151.5 AND kolestrol >311 AND kolestrol > 237.5 THEN class = SICK

39. R39: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117 AND kolestrol > 151.5 AND kolestrol >311 AND kolestrol ≤ 237.5 AND flaurosofy = 0 AND umur > 46 THEN class = HEALTHY
40. R40: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117 AND kolestrol > 151.5 AND kolestrol >311 AND kolestrol ≤ 237.5 AND flaurosofy = 0 AND umur ≤ 46 THEN class = SICK
41. R341: IF denyut_jantung = reversable_defect AND oldpeak ≤ 0.7 AND tekanan_darah > 122 AND tekanan_jantung > 117 AND kolestrol > 151.5 AND kolestrol >311 AND kolestrol ≤ 237.5 AND flaurosofy = 1 THEN class = SICK
42. R42: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = asympt THEN class = SICK
43. R43: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = atyp_angina AND jenis_kelamin = female THEN class = HEALTHY
44. R44: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = atyp_angina AND jenis_kelamin = male THEN class = SICK
45. R45: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = non_anginal AND oldpeak >1.9 THEN class = SICK
46. R46: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = non_anginal AND oldpeak ≤ 1.9 AND tekanan_jantung >150.5 THEN class = HEALTHY
47. R47: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = non_anginal AND oldpeak ≤ 1.9

AND tekanan_jantung ≤150.5 flaurosofy = 0 THEN class = SICK

48. R48: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = non_anginal AND oldpeak ≤ 1.9 AND tekanan_jantung ≤150.5 flaurosofy = 1 THEN class = SICK
49. R49: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = non_anginal AND oldpeak ≤ 1.9 AND tekanan_jantung ≤150.5 flaurosofy = 3 THEN class = HEALTHY
50. R50: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = typ_angina AND kolestrol > 215 AND umur > 48.5 THEN class = HEALTHY
51. R51: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = typ_angina AND kolestrol > 215 AND umur ≤ 48.5 THEN class = SICK
52. R52: IF denyut_jantung = reversable_defect AND oldpeak > 0.7 AND jenis_sakit dada = typ_angina AND kolestrol ≤ 215 THEN class = HEALTHY
53. R53: IF denyut_jantung = normal AND falurosofy =2 AND angina_induksi = no AND oldpeak >2.3 THEN class = SICK
54. R54: IF denyut_jantung = normal AND falurosofy =2 AND angina_induksi = no AND oldpeak ≤2.3 AND kolestrol > 311 THEN class = SICK
55. R55: IF denyut_jantung = normal AND falurosofy =2 AND angina_induksi = no AND oldpeak ≤2.3 AND kolestrol ≤ 311 THEN class = HEALTHY
56. R56: IF denyut_jantung = normal AND falurosofy =2 AND angina_induksi = yes THEN class = SICK
57. R57: IF denyut_jantung = normal AND falurosofy =3 AND tekanan_jantung > 167.5 THEN class = HEALTHY
58. R57: IF denyut_jantung = normal AND falurosofy =3 AND tekanan_jantung ≤ 167.5 THEN class = SICK

Lampiran 4. Hasil perhitungan seluruh atribut, gambar model pohon keputusan dan rule dari 10 model Algoritma C4.5 berbasis Adaboost

Tabel Informasi Gain untuk Algoritma C4.5 Berbasis Adaboost

Model 1 [w = 4.538]

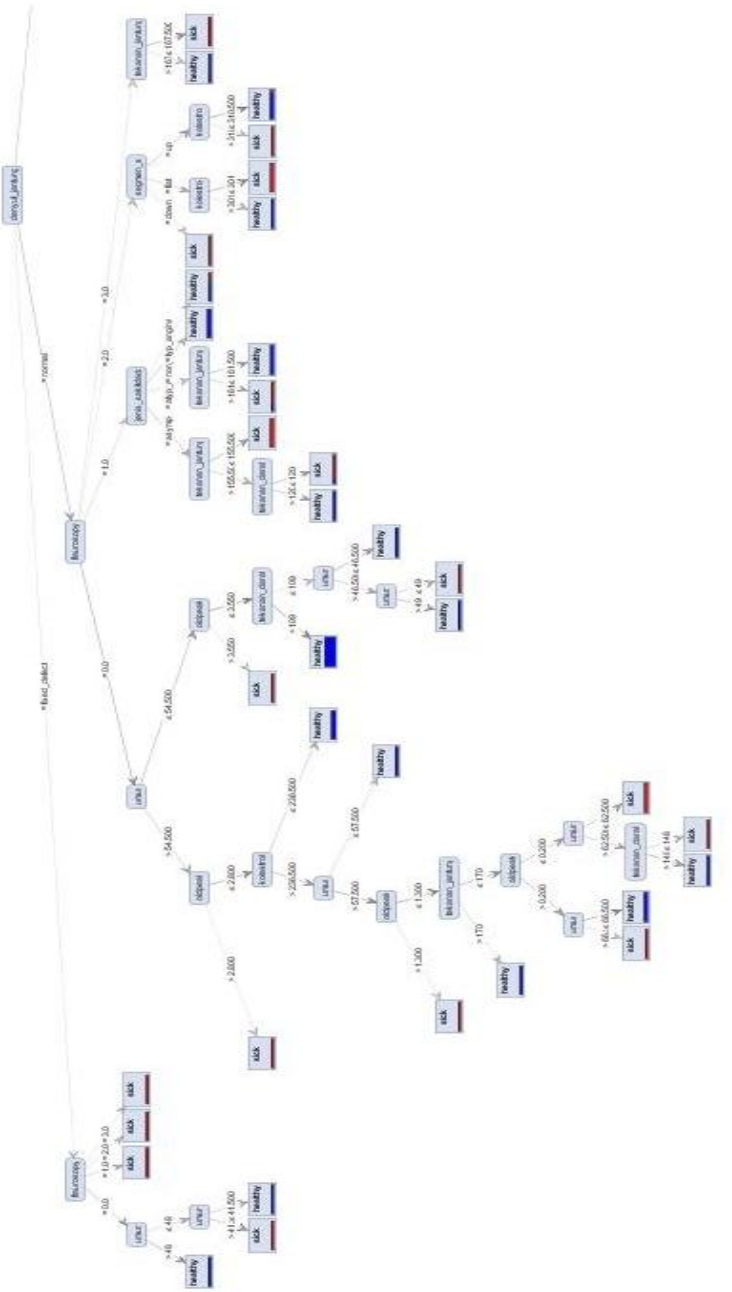
Candidate split	Kasus	SICK	HEALTHY	Entropy	Gian
Denyut_jantung					
= fixed_defect	32	20	12	0,0539	
= normal	315	69	246	0,4213	
= reversable_defect	220	168	52	0,3061	
	334	85	249	0,4820	
	123	82	41	0,1992	
Flaurosofy	71	57	14	0,0897	
= 0	39	33	6	0,0426	
= 1					
= 2	311	159	152	0,9817	0,0119
= 3	256	98	158		
Kolestrol	84	39	45	0,9936	0,0001
> 236.500	483	218	265		
≤ 236.500	40	21	19	0,9692	0,0245
>301	514	229	285		
≤ 301	309	157	152	0,9832	0,0105
> 310,500	258	100	158		
≤ 310,500	377	186	191	0,9843	0,0094
> 237,500	190	71	119		
≤ 237,500	102	49	53	0,9932	0,0005
> 225	465	208	257		
≤ 225	202	108	94	0,9830	0,0107
> 293	365	149	216		
≤ 293	145	83	62	0,9795	0,0142
> 262	422	174	248		
≤ 262	355	177	178	0,9836	0,0101
	212	80	132		

> 275,500					
≤ 275,500	103	56	47	0,9884	0,0053
> 229	464	201	263		
≤ 229	531	249	282	0,9824	0,0113
Tekanan_darah	36	8	28		
> 148	383	187	196	0,9862	0,0075
≤ 148	184	70	114		
> 109	376	185	191	0,9850	0,0087
≤ 109	191	72	119		
> 120	376	185	191	0,9850	0,0087
≤ 120	191	71	119		
> 122					
≤ 122	412	212	200	0,9638	0,0229
> 122,500	155	110	110		
≤ 122,500	516	243	273	0,9841	0,0096
Umur	51	14	37		
> 48	233	138	95	0,9542	0,0394
≤ 48	334	119	215		
> 41,500	50	24	26	0,9935	0,0002
≤ 41,500	517	233	284		
> 57,500	114	56	58	0,9926	0,0011
≤ 57,500	453	201	252		
> 66,500	296	174	122	0,9352	0,0585
≤ 66,500	271	83	188		
> 62,500	436	223	213	0,9595	0,0342
≤ 62,500	131	34	97		
> 54,500	402	208	194	0,9637	0,0299
≤ 54,500	165	49	116		
> 46,500	500	241	259	0,9747	0,0190
≤ 46,500	67	16	51		
> 49	412	212	200	0,9638	0,0299
≤ 49	155	45	110		
> 42					
≤ 42	124	31	93	0,9585	0,0352
> 48,500	443	226	217		

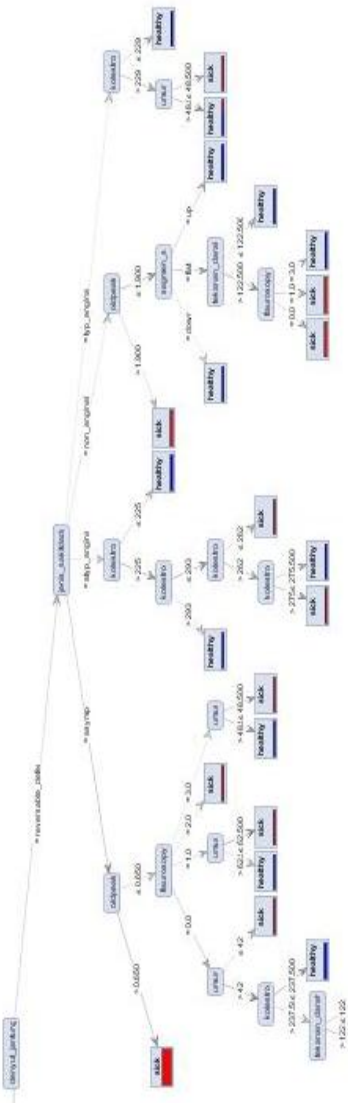
≤ 48,500	279	83	196	0,9240	0,0696
Tekanan_jantung	288	174	114		
> 170	211	57	154	0,9341	0,0595
≤ 170	356	200	156		
> 155,500	153	41	112	0,9554	0,0383
≤ 155,500	414	216	198		
> 161,500					
≤ 161,500	48	42	6	0,9418	0,0518
> 167,500	519	215	304		
≤ 167,500	200	138	62	0,9034	0,0903
Oldpeak	367	119	248		
> 2,800	352	201	151	0,9255	0,0682
≤ 2,800	215	56	159		
> 1,300	24	22	2	0,9627	0,0310
≤ 1,300	543	235	308		
> 0,200	294	187	107	0,8859	0,1078
≤ 0,200	273	70	203		
> 3,550	111	93	18	0,8831	0,1106
≤ 3,550	456	164	292		
> 0,650					
≤ 0,650	39	22	17	0,0680	
> 1,900	260	168	92	0,4299	
≤ 1,900	268	67	201	0,3835	
Segment_st					
= down	23	7	16	0,0360	
= flat	71	15	56	0,0932	
= up	162	35	127	0,2151	
Jenis_sakit dada	311	200	111	0,5156	
= typ_angina					
= atyp_angina					
= non_anginal					
= asympt					

Sehingga pohon keputusan dari hasil perhitungan diatas dapat digambarkan pada Gambar dibawah ini:

Gambar Model Pohon Keputusan Algoritma C4.5 berbasis Adaboost Model 1 [w = 4.538]



Gambar Model Pohon Keputusan Algoritma C4.5 berbasis Adaboost Model 1 [w = 4.538] (2)



Adapun rule yang didapat dari Gambar diatas adalah:

1. R1: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND umur >48 THEN class = HEALTHY
2. R2: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND umur ≤48 AND umur >41,5 THEN class = SICK
3. R3: IF denyut_jantung = fixed_defect AND flaurosopy = 0 AND umur ≤48 AND umur ≤41,5 THEN class = HEALTHY
4. R4: IF denyut_jantung = fixed_defect AND flaurosopy = 1 THEN class = SICK
5. R5: IF denyut_jantung = fixed_defect AND flaurosopy = 2 THEN class = SICK
6. R6: IF denyut_jantung = fixed_defect AND flaurosopy = 3 THEN class = HEALTHY
7. R7: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 THEN class = SICK
8. R8: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 THEN class = SICK
9. R9: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 AND tekanan_jantung >170 THEN class = HEALTHY
10. R10: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak ≤1,3 AND tekanan_jantung ≤170 AND oldpeak >0,2 AND umur >66,5 THEN class = SICK
11. R11: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 AND tekanan_jantung ≤170 AND oldpeak >0,2 AND umur ≤66,5 THEN class = HEALTHY

12. R12: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak ≤2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 AND tekanan_jantung ≤170 AND oldpeak ≤0,2 AND umur >62,5 AND tekanan_darah >148 THEN class = HEALTHY
13. R13: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak ≤2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 AND tekanan_jantung ≤170 AND oldpeak ≤0,2 AND umur >62,5 AND tekanan_darah ≤148 THEN class = SICK
14. R14: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak ≤2,8 AND kolestrol >236,5 AND umur > 57,5 AND oldpeak >1,3 AND tekanan_jantung ≤170 AND oldpeak ≤0,2 AND umur ≤62,5 THEN class = SICK
15. R15: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol >236,5 AND umur ≤57,5 THEN class = HEALTHY
16. R16: IF denyut_jantung = normal AND flaurosopy = 0 AND umur >54,5 AND oldpeak >2,8 AND kolestrol ≤236,5 THEN class = HEALTHY
17. R17: IF denyut_jantung = normal AND flaurosopy = 0 AND umur ≤54,5 AND oldpeak >3,55 THEN class = SICK
18. R18: IF denyut_jantung = normal AND flaurosopy = 0 AND umur ≤54,5 AND oldpeak ≤3,55 AND tekanan_darah >109 THEN class = HEALTHY
19. R19: IF denyut_jantung = normal AND flaurosopy = 0 AND umur ≤54,5 AND oldpeak ≤3,55 AND tekanan_darah ≤109 AND umur >46,5 AND umur > 49 THEN class = HEALTHY
20. R20: IF denyut_jantung = normal AND flaurosopy = 0 AND umur ≤54,5 AND oldpeak ≤3,55 AND tekanan_darah ≤109 AND umur >46,5 AND umur ≤49 THEN class = SICK

21. R21: IF denyut_jantung = normal AND flaurosopy = 0 AND umur ≤54,5 AND oldpeak ≤3,55 AND tekanan_darah ≤109 AND umur ≤46,5 THEN class = HEALTHY
22. R22: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =asympt AND tekanan_jantung >155,5 AND tekanan_darah >120 THEN class = HEALTHY
23. R23: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =asympt AND tekanan_jantung >155,5 AND tekanan_darah ≤120 THEN class = SICK
24. R24: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =asympt AND tekanan_jantung ≤155,5 AND THEN class = SICK
25. R25: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =atyp_angina AND tekanan_jantung >161,5 AND THEN class = SICK
26. R26: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =atyp_angina AND tekanan_jantung ≤161,5 AND THEN class = HEALTHY
27. R27: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =non_anginal THEN class = HEALTHY
28. R28: IF denyut_jantung = normal AND flaurosopy = 1 AND jenis_sakitdada =typ_angina THEN class = HEALTHY
29. R29: IF denyut_jantung = normal AND flaurosopy = 2 AND segment_st = down THEN class = SICK
30. R30: IF denyut_jantung = normal AND flaurosopy = 2 AND segment_st = flat AND kolestrol >301 THEN class = HEALTHY
31. R31: IF denyut_jantung = normal AND flaurosopy = 2 AND segment_st = flat AND kolestrol ≤301 THEN class = SICK
32. R32: IF denyut_jantung = normal AND flaurosopy = 2 AND segment_st = up AND kolestrol >310,5 THEN class = SICK

33. R33: IF denyut_jantung = normal AND flaurosopy = 2 AND segment_st = up AND kolestrol ≤310,5 THEN class = HEALTHY
34. R34: IF denyut_jantung = normal AND flaurosopy = 3 AND tekanan_jantung >167,5 THEN class = HEALTHY
35. R35: IF denyut_jantung = normal AND flaurosopy = 3 AND tekanan_jantung ≤167,5 THEN class = SICK
36. R36: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak >0,65 THEN class = SICK
37. R37: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 0 AND umur >42 AND kolestrol >237,5 AND tekanan_darah >122 THEN class = SICK
38. R38: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 0 AND umur >42 AND kolestrol >237,5 AND tekanan_darah ≤122 THEN class = HEALTHY
39. R39: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 0 AND umur >42 AND kolestrol ≤237,5 THEN class = HEALTHY
40. R40: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 0 AND umur ≤42 THEN class = SICK
41. R41: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 1 AND umur >62 THEN class = HEALTHY
42. R42: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 1 AND umur ≤62 THEN class = SICK

43. R43: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 2 THEN class = SICK
44. R44: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 3 AND umur >48,5 THEN class = HEALTHY
45. R45: IF denyut_jantung = reversible_defect AND jenis_sakitdada = asympt AND oldpeak ≤0,65 AND flaurosofy = 3 AND umur ≤48,5 THEN class = SICK
46. R46: IF denyut_jantung = reversible_defect AND jenis_sakitdada = atyp_angina AND kolestrol >225 AND kolestrol >293 THEN class = HEALTHY
47. R47: IF denyut_jantung = reversible_defect AND jenis_sakitdada = atyp_angina AND kolestrol >225 AND kolestrol ≤293 AND kolestrol >265 AND kolestrol >275,5 THEN class = SICK
48. R48: IF denyut_jantung = reversible_defect AND jenis_sakitdada = atyp_angina AND kolestrol >225 AND kolestrol ≤293 AND kolestrol >265 AND kolestrol ≤275,5 THEN class = HEALTHY
49. R49: IF denyut_jantung = reversible_defect AND jenis_sakitdada = atyp_angina AND kolestrol >225 AND kolestrol ≤293 AND kolestrol ≤265 THEN class = SICK
50. R50: IF denyut_jantung = reversible_defect AND jenis_sakitdada = atyp_angina AND kolestrol ≤225 THEN class = HEALTHY
51. R51: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak >1,9 THEN class = SICK
52. R52: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = down THEN class = HEALTHY

53. R53: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = flat AND tekanan_darah >122.5 AND falurosofy = 0 THEN class = SICK
54. R54: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = flat AND tekanan_darah >122.5 AND falurosofy = 1 THEN class = SICK
55. R55: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = flat AND tekanan_darah >122.5 AND falurosofy = 3 THEN class = HEALTHY
56. R56: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = flat AND tekanan_darah ≤122.5 THEN class = HEALTHY
57. R57: IF denyut_jantung = reversible_defect AND jenis_sakitdada = non_anginal AND oldpeak ≤1,9 AND segment_st = up THEN class = HEALTHY
58. R58: IF denyut_jantung = reversible_defect AND jenis_sakitdada = typ_angina AND kolestrol >229 AND umur >48,5 THEN class = HEALTHY
59. R59: IF denyut_jantung = reversible_defect AND jenis_sakitdada = typ_angina AND kolestrol >229 AND umur ≤48,5 THEN class = SICK
60. R56: IF denyut_jantung = reversible_defect AND jenis_sakitdada = typ_angina AND kolestrol ≤229 THEN class = HEALTHY

SINOPSIS

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di Indonesia, penyakit jantung sangat berbahaya karena dapat menimbulkan serangan jantung dan kematian mendadak. Oleh karena perlu kita harus waspada terhadap gejala-gejalanya.

Dengan adanya suatu prediksi terhadap penyakit jantung yang tepat, maka dapat dijadikan suatu informasi untuk mengenali, memahami dan mewaspadai terjadi serangan jantung. Atribut yang digunakan sebagai acuan timbulnya gejala yaitu; umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, kecepatan detak jantung, angina induksi, oldpeak, sengemt_st, floirosopy dan denyut jantung.

Algoritma *Decision Tree C4.5* berbasis *Adaboost* merupakan metode data mining yang baik dan banyak digunakan untuk membuat suatu prediksi terutama penyakit jantung, karena mudah untuk dipahami melalui hasil pola *decision tre*e atau pohon keputusannya dan juga nilai menghasilkan akurasi performancenya yang baik.

Maka dengan buku monograf ini yang berjudul: "prediksi penyakit jantung dengan menggunakan algoritma *C4.5* berbasis *Adaboost"* memberikan wawasan dan ilmu pengetahuan kepada pembaca baik dosen, mahasiswa maupun umum.

BIODATA PENULIS

Abdul Rohman S.Pd., M.Kom, Kelahiran Bogor, tanggal 15 Juni 1982, Lulus S1 Teknologi Pendidikan UNNES pada tahun 2004 dan S2 Teknik Informatika UDINUS pada tahun 2013, Saat ini aktif sebagai Dosen di Program Studi S1 Teknik Informatika Universitas Ngudi Waluyo Ungaran dan Instruktur di e-guru.id. Abdul Rohman aktif menulis dan membuat konten di situs blog dan Channel YouTube dengan bidang IT dan Pembelajaran.

MONOGRAF

PREDIKSI
Penyakit Jantung

MENGGUNAKAN ALGORITMA C4.5 BERBASIS ADABOOST

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di Indonesia, penyakit jantung sangat berbahaya karena dapat menimbulkan serangan jantung dan kematian mendadak. Oleh karena itu, kita harus waspada terhadap gejala-gejalanya.

Dengan adanya suatu prediksi terhadap penyakit jantung yang tepat, maka dapat dijadikan suatu informasi untuk mengenali, memahami dan mewaspadai terjadi serangan jantung. Atribut yang digunakan sebagai acuan timbulnya gejala, yaitu umur, jenis kelamin, jenis sakit dada, tekanan darah, kolestrol, kadar gula, elektrokardiografi, kecepatan detak jantung, angina induksi, oldpeak, sengemt_st, floirosopy dan denyut jantung.

Dengan permasalahan-permasalahan di atas, di buku ini penulis ingin berbagi informasi terkait prediksi penyakit jantung menggunakan algoritma *C4.5* berbasis *Adaboost*. Semoga buku ini bisa bermanfaat bagi para pembaca.

PENERBIT LAKEISHA
Jl. Jatinom Boyolali,
Srikaton, Rt.003, Rw.001,
Pucangmiliran, Tulung,
Klaten, Jateng, Indonesia 57482
Email :penerbit_lakeisha@yahoo.com
HP/WA :08989880852
Website : http://www.penerbitlakeisha.com/

ISBN 978-623-420-004-1