

**PENERAPAN METODE DECISION TREE DENGAN
ALGORITMA C4.5 UNTUK KLASIFIKASI
PENYAKIT JANTUNG**

PROPOSAL SKRIPSI



Disusun Oleh :
Rama Triyandika
8020190046

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir
Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DINAMIKA BANGSA
2022**

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Penerapan Metode Decision Tree Dengan Algoritma C4.5
Untuk Klasifikasi Penyakit jantung.

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

- a. Nama Lengkap : Rama Triyandika
- b. NIM : 8020190046
- c. Jenis Kelamin : Laki-Laki
- d. Tempat/Tgl Lahir : Jambi/24 Desember 2001
- e. Alamat : Jl. Sukasari, RT.39, No.30,
Kec. Jambi selatan, Kebun Kopi
- f. No. Telepon : 085357500896
- g. Email : ramatriandika3@gmail.com

ABSTRAK

Penyakit Jantung (Heart Disease) Merupakan kondisi ketika jantung mengalami gangguan. Bentuk gangguan itu sendiri bermacam-macam, bisa berupa gangguan pada pembuluh darah jantung, katup jantung, atau otot jantung. penyakit jantung memiliki angka kematian yang cukup tinggi dan merupakan penyakit yang ditakuti masyarakat. Dengan banyaknya kasus yang telah terjadi maka diperlukan algoritma untuk mengklasifikasikan apakah seseorang mengidap penyakit jantung atau tidak. Penelitian ini menggunakan Algoritma C4.5 sebagai metode dalam klasifikasi penyakit jantung.

Kata Kunci : Klasifikasi, *Data Mining*, *Algoritma C4.5*, *Penyakit Jantung*

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung” dengan sebaik-baiknya.

Penulis menyadari bahwa dalam penulisan proposal skripsi ini banyak mengalami kendala dan kekurangan, namun berkat bantuan, bimbingan, dan petunjuk dari banyak pihak sehingga kendala-kendala yang dihadapi tersebut dapat diatasi dan penulis dapat menyelesaikan proposal skripsi ini dengan baik.

Penulis berharap agar proposal skripsi ini dapat bermanfaat bagi kita semua. Akhir kata penulis mengucapkan terima kasih.

Jambi, September 2022

Penulis

DAFTAR ISI

	Halaman
IDENTITAS PROPOSAL PENELITIAN	ii
ABSTRAK.....	iii
KATA PENGANTAR	iv
DAFTAR ISI.....	v
BAB I PENDAHULUAN	1
1.1. LATAR BELAKANG.....	1
1.2. RUMUSAN MASALAH	2
1.3. BATASAN MASALAH	3
1.4. TUJUAN DAN MANFAAT PENELITIAN.....	3
1.4.1 Tujuan Penelitian.....	3
1.4.2 Manfaat Penelitian.....	3
BAB II LANDASAN TEORI	4
2.1. DEFINISI PENYAKIT	4
2.2. DATA MINING	4
2.2.1. Tahapan Proses dalam <i>Data Mining</i>	5
2.2.2. Pengelompokan <i>Data Mining</i>	6
2.3. KLASIFIKASI	8
2.4. <i>DECISION TREE (Algoritma C4.5)</i>	8
2.5. PENELITIAN SEJENIS	10
BAB III METODOLOGI PENELITIAN	16
3.1. KERANGKA KERJA PENELITIAN	16
3.2. ANALISA TEKNIK <i>ALGORITMA C4.5</i>	18
3.3. METODE PENELITIAN	18
3.4. ALAT PENELITIAN	18
3.5. JADWAL PENELITIAN	19
DAFTAR PUSTAKA	20

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Dengan pesatnya perkembangan teknologi, kebutuhan akan informasi yang akurat menjadi sangat penting dalam kehidupan kita sehari-hari dan di masa mendatang. Metode tradisional untuk menganalisis informasi yang ada, sebagian besar informasi tidak dapat diperbaiki. Oleh karena itu, informasi dapat diperbaiki atau diproses menghasilkan suatu pengetahuan dengan istilah *data mining* [1]. *Data mining* dapat diterapkan di berbagai bidang, salah satunya di bidang kesehatan. Pengambilan keputusan berdasarkan data dan informasi yang akurat akan menghasilkan keputusan dan prediksi penyakit yang menjadi target [2].

Pada 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 dokter. Pengambilan keputusan atas dasar data dan informasi yang akurat akan menghasilkan keputusan dan prediksi penyakit menjadi tepat sasaran.

Penyakit jantung adalah kondisi ketika jantung mengalami gangguan. Bentuk gangguan itu sendiri bermacam-macam, bisa berupa gangguan pada pembuluh darah jantung, katup jantung, atau otot jantung. Penyakit jantung merupakan penyakit yang memiliki angka kematian yang cukup tinggi dan merupakan penyakit yang ditakuti masyarakat. Oleh karena itu, klasifikasi data mining diperlukan untuk membantu dokter dalam pengambilan keputusan yang cepat dan akurat.

Banyak penelitian yang memprediksi penyakit jantung dengan teknik klasifikasi *data mining*, diantaranya dilakukan oleh [3] menggunakan *Algoritma Back Propagation Neural Network* dengan Metode *Adaboost*. Hasil penelitian bahwa nilai akurasi hasil prediksi menggunakan algoritma BPNN adalah 96,65 % dan algoritma BPNN dengan metode *Adaboost* menjadi 99,29 %. Pada penelitian [4] menggunakan metode *Extreme Learning Machine*

(ELM) dan *Algoritma C4.5* adalah *Algoritma C4.5* mempunyai tingkat akurasi yang lebih tinggi dibandingkan dengan *ELM* dikarenakan *Algoritma C4.5* dijabarkan berdasarkan nilai gain tertinggi yang dijadikan nodes dan hasil akhirnya dalam bentuk leaves yang berisikan indikator penentu seseorang penyakit jantung koroner atau tidak.

Algoritma C4.5 adalah algoritma yang pada umumnya digunakan guna untuk membentuk *decision tree*. *Decision tree* adalah salah satu metode klarifikasi yang paling populer karena mudah diinterpretasi oleh manusia. Kelebihan *algoritma C4.5* adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani atribut yang bertipe diskret atau numerik.

Berdasarkan latar belakang yang telah dikemukakan, maka judul yang diambil dalam skripsi ini adalah “***Penerapan Metode Decision Tree Dengan Algoritma C4.5 Untuk Klasifikasi Penyakit Jantung***”.

1.2. RUMUSAN MASALAH

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *Algoritma C4.5* dalam mengklasifikasi penyakit jantung?
2. Seberapa besar tingkat akurasi *Algoritma C4.5* apabila digunakan untuk mengklasifikasi penyakit jantung?

1.3. BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu:

1. Data yang diperlukan adalah dataset Penyakit Jantung.
2. Variabel yang digunakan dalam penelitian ini adalah *Age*, *Sex*, *ChestPainType*, *RestingBP*, *Cholesterol*, *FastingBP*, *RestingECG*, *MaxHR*, *ExerciseAngina*, *OldPeak*, *ST_Slope*, yang memberikan informasi mengenai penyakit jantung.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Mengklasifikasi Penyakit Jantung
2. Mendapatkan akurasi yang tepat untuk melakukan klasifikasi Penyakit Jantung dengan menggunakan *Algoritma C4.5*.
3. Mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi Penyakit jantung.

1.4.2. Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *data mining* khususnya untuk klasifikasi sebuah *dataset*.
2. Menambah wawasan bagi penulis dalam pengklasifikasian *dataset* Penyakit Jantung.

BAB II

LANDASAN TEORI

2.1. DEFINISI PENYAKIT

Penyakit dikenal sebagai sebagai kondisi medis yang dihubungkan dengan gejala dan tanda klinis tertentu. Penyakit adalah suatu kondisi abnormal tertentu yang secara negatif memengaruhi struktur atau fungsi sebagian atau seluruh tubuh makhluk hidup.

2.2. DATA MINING

Data mining sering juga disebut *Knowledge Discovery in Database*, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan [5].

Menurut Turban et al dalam [6] *data mining* adalah :

“*Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar”.

Selain beberapa definisi diatas beberapa definisi juga diberikan seperti,

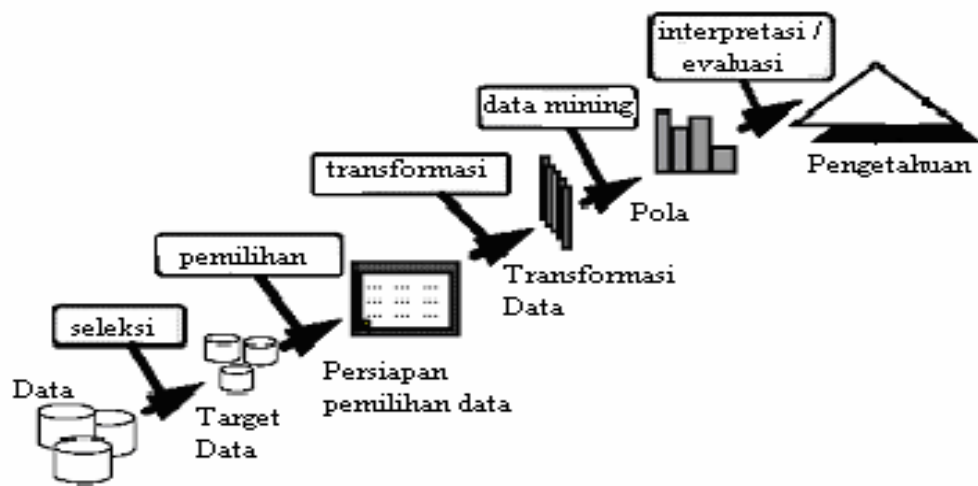
Pramudiono dalam [6] mendefenisikan “*Data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya”.

Berdasarkan definisi – definisi yang telah disampaikan, hal penting yang terkait dengan *data mining* adalah :

1. *Data mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan di proses berupa data yang sangat besar. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

2.2.1. Tahapan Proses dalam *Data Mining*

Ada beberapa tahapan dalam proses *data mining*. Diagram dibawah ini menggambarkan beberapa tahap/proses yang berlangsung dalam *data mining*. Fase awal dimulai dari data sumber dan berakhir dengan adanya informasi yang dihasilkan dari beberapa tahapan, yaitu :



Gambar 2.1. Fase – Fase dalam *Data Mining*

Tahapan dalam proses *data mining* dapat dijelaskan sebagai berikut :

1. Seleksi data
Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *data mining* dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.
2. *Pre-processing/Cleaning* (pemilihan data)
Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

3. Transformasi

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*.

4. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Pemilihan metode dan algoritma sesuai dengan kebutuhan dan tujuan.

5. Interpretasi/Evaluasi

Pola informasi yang dihasilkan dari informasi *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan

2.2.2. Pengelompokan *Data Mining*

Pengelompokan *data mining* dapat diklasifikasikan berdasarkan fungsi yang dilakukan atau berdasarkan jenis aplikasi yang menggunakannya [6].

1. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

Contoh lain klasifikasi dalam bisnis dan penelitian adalah :

- a. Menentukan apakah suatu transaksi kartu kredit merupakan transaksi yang curang atau bukan.
- b. Memperkirakan apakah suatu pengajuan hipotek oleh nasabah merupakan suatu kredit yang baik atau buruk.
- c. Mendiagnosa penyakit seorang pasien untuk mendapatkan termasuk kategori apa.

2. Pengklusteran (*Clustering*)

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan

suatu dengan yang lainnya dan memiliki ketidakmiripan dengan *record* dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan (homogen), yang mana kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

3. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Contoh asosiasi dalam bisnis dan penelitian adalah :

- a. Meneliti jumlah pelanggan dari perusahaan telekomunikasi seluler yang diharapkan untuk memberikan respon positif terhadap penawaran *upgrade* layanan yang diberikan.
- b. Menemukan barang dalam supermarket yang dibeli secara bersamaan dan barang yang tidak pernah dibeli bersamaan.

4. Deskripsi

Terkadang analis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

5. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai

contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan, dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

6. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

2.3. KLASIFIKASI

Klasifikasi adalah tugas pembelajaran sebuah fungsi target f yang memetakan setiap himpunan atribut x ke salah satu label *class* y yang telah didefinisikan sebelumnya [7].

Klasifikasi dapat juga diartikan suatu proses untuk menemukan suatu model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep dengan tujuan dapat menggunakan model untuk memprediksi kelas objek yang label *class*-nya tidak diketahui.

2.4. DECISION TREE (Algoritma C4.5)

Pohon keputusan (*Decision Tree*) merupakan salah satu metode data mining yang banyak diterapkan sebagai solusi untuk mengklasifikasikan masalah. *Decision tree* adalah suatu struktur yang berbentuk seperti pohon dimana setiap simpul (*node*) pohon mewakili atribut yang telah diuji. Setiap cabang mewakili hasil yang telah diuji dan simpul daun (*leaf node*) mewakili kelas atau distribusi kelas [8].

Algoritma C4.5 merupakan sebuah algoritma yang digunakan untuk membangun decision tree (pohon keputusan). *Algoritma C4.5* dan pohon keputusan merupakan dua model yang tak terpisahkan. *Algoritma C4.5* merupakan salah satu algoritma klasifikasi yang kuat dan cukup banyak digunakan atau di implementasikan untuk pengklasifikasian dalam berbagai hal. Dari beberapa algoritma yang dapat digunakan untuk membuat pohon keputusan (decision tree) yaitu diantaranya *algoritma C4.5*.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 [9] yaitu:

- a. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
- b. 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:

$$Entropy(i) = \sum_{j=1}^m f(i, j) \cdot \log_2 f(i, j)$$

- c. Hitung nilai *Gain* dengan rumus:

$$gain = - \sum_{i=1}^p \frac{n_i}{n} \cdot IE(i)$$

- d. Selanjutnya menghitung *Gain ratio*:

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

- e. Ulangi langkah ke-2 hingga semua record terpartisi. Proses partisi pohon keputusan akan berhenti disaat :
 - 1.) Semua tupel dalam record dalam simpul N mendapat kelas yang sama,
 - 2.) Tidak ada atribut dalam record yang dipartisi lagi,
 - 3.) Tidak ada record dalam cabang yang kosong.

2.5. PENELITIAN SEJENIS

Tabel 2.1. Penelitian Sejenis

No.	Judul, Penulis dan Tahun	Metode	Atribut	Akurasi
1	“Perbandingan Metode <i>Algoritma C4.5</i> Dan <i>Extreme Learning Machine</i> Untuk Mendiagnosis Penyakit Jantung Koroner.” Jefri Junifer Pangaribuan, Cathlin Tedja , Sentosa Wibowo, 2019 [4].	<i>Algoritma C4.5, Dan Extreme Learning Machine</i>	Umur, Jenis Kelamin, <i>Chest Pain Type, Resting Blood Pressure, Serum Cholestrol, Fasting Blood Sugar, Resting Electrocardiographic Results, Maximum Heart Rate Achieved, Exercise Induced Angina, ST Depression Induced By Exercise Relative To Rest, The Slope Of The Peak Exercise ST Segment, Number Of Major Vessels, Dan Thal.</i>	Hasil yang diperoleh dari percobaan dan analisis diagnosis penyakit jantung koroner dengan menggunakan <i>Aloritma C4.5</i> sebesar 93,33%. Lebih tinggi dibandingkan metode <i>Extreme Learning Machine (ELM)</i> yaitu sebesar 73,33%, lebih
2	“Peningkatan Akurasi Klasifikasi <i>Algoritma C4.5</i> Menggunakan Teknik Bagging Pada Diagnosis Penyakit Jantung.” Erwin Prasetyo, Budi Prasetyo. 2019 [10].	<i>Algoritma C4.5</i>	Usia, Sex, Cp, Chol, Fbs, Exang, Slope, Ca, Oldpeak, Num	Klasifikasi menggunakan <i>Algoritma C4.5</i> dalam memprediksi penyakit jantung menghasilkan akurasi sebesar 72,98%. Dari hasil ini ditingkatkan dengan menerapkan teknik bagging pada <i>Algoritma C4.5</i> , menghasilkan akurasi sebesar

				81,84%.
3	“Diagnosa Prediksi Penyakit Jantung Dengan Model <i>Algoritma Naïve Bayes</i> Dan <i>Algoritma C4.5</i>.” Tri Retnasari, Eva Rahmawati, 2017 [11].	<i>Algoritma Naïve Bayes</i> Dan <i>Algoritma C4.5</i>	Umur, Jenis Kelamin, Jenis Sakit Dada, Tekanan Darah, Kolesterol, Kadar Gula, Elektrokardiografi, Tekanan Darah, Angina Induksi, <i>Oldpeak</i>, <i>Segmen_St</i>, <i>Fluoroscopy</i>, Denyut Jantung.	Nilai accuracy dan AUC <i>Naïve Bayes</i> sebesar 86,67% lebih tinggi dibandingkan dengan <i>Algoritma C 4.5</i>. sebesar 83,70%. Terlihat bahwa nilai accuracy dan AUC <i>Naïve Bayes</i> lebih tinggi dibandingkan C 4.5. Penerapan <i>Naïve Bayes</i> untuk prediksi penyakit jantung menghasilkan selisih nilai akurasi sebesar 2.97%.
4	“Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan <i>Algoritma C4.5</i>.” Elin Nurlia, Ultach Enri, 2021 [12].		<i>Age, Sex, Serum_Creatinine, Time, Anemia, Diabetes, High_Blood_Pressure, Smoking, Creatinine Phosphokinase, Ejection_Fraction, Platelets, Serum_Sodium</i>	Dengan penambahan seleksi fitur forward selection mampu meningkatkan performa <i>algoritma C4.5</i> dalam melakukan klasifikasi dengan akurasi sebesar 84,29% dan nilai AUC 0,785 yang termasuk kategori <i>fair classification..</i> menghasilkan
5	“Komparasi	<i>Random</i>	<i>Age, Sex,</i>	menghasilkan

	Algoritma Klasifikasi Data Mining untuk Memprediksi Penyakit Jantung”. Dimas Anugrah Firdlous. 2022 [13].	<i>Forest, Naïve Bayes, Decision Tree, Support vector machine</i>	<i>ChestPaintype, RestingBP, Cholesterol, RestingBS, Resting ECG, MaxHR, ExerciseAngina, Oldpeak, ST_Slope, HeartDisease</i>	algoritma random forest sebagai algoritma dengan ukuran akurasi tertinggi yang memiliki akurasi sebesar 85,9% kemudian di urutan kedua menggunakan <i>naive bayes</i> dengan akurasi 84,4%, kemudian menggunakan <i>decision tree</i> dengan akurasi sebesar 72,1%, dan terakhir adalah menggunakan <i>support vector machine</i> dengan akurasi 68,7%.
6	“Penerapan Data Mining Menggunakan <i>Algoritma C4.5</i> Dalam Prediksi Penyakit Angin Duduk”. Salman Alfaridzi, Agung Nugroho, Muhammad Rizki Sani. 2021 [14].	<i>Algoritma C4.5</i>	<i>Status, Age, Smoke, Cig, Hyper, Fam ang, Fam myo, Fam Stroke</i>	Pengujian yang dilakukan menggunakan Particle Swarm Optimization (PSO) pada metode <i>Algoritma C4.5</i> dalam menganalisis prediksi penyakit angin duduk dapat meningkatkan nilai Accuracy, Precision, dan Recall. Nilai

				Accuracy yang didapatkan meningkat sebanyak 7,5% dari 76,50% menjadi 84,00%, nilai Precision yang didapatkan meningkat sebanyak 7,64% dari 80,50% menjadi 88,14%, dan nilai Recall yang didapatkan meningkat sebanyak 9% dari 72,00% menjadi 81,00%.
7	<i>“Comparison of Data Mining Methods Using C4.5 Algorithm and Naive Bayes in Predicting Heart Disease. Rino”. 2021 [15].</i>	<i>Algoritma C4.5, Naïve Bayes</i>	<i>Sex, Age, CSk, CPD PS, PHyp, Dbts, TCh SBP, DBP, BMI, HR</i>	The expert system using the Naïve Bayes method has an accuracy of 83.17% while the C4.5 algorithm only has an accuracy of 81.77%, so the Naïve Bayes algorithm has a superior accuracy value.
8	<i>“An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease.” Ahmed Abba Haruna, L. J. Muhammad, B. Z. Yahaya,</i>	<i>Algoritma c4.5</i>	<i>Age, Blood Pressure, Glucose, Cholesterol, Triglyceride, High Density of Lipoprotein, Low Density of Lipoprotein, Creatinine, Body mass, Heart rate, Chest pain.</i>	The improved C4.5 algorithm has shown a slightly better performance for the diagnosis of CAD dataset with an overall accuracy of 97.23%, 97.03% specificity, and 96.39%

	E. J. Garba, N. D. Oye, Low Tan Jung. 2019 [16].			sensitivity. While <i>C4.5 algorithm</i> has an overall accuracy of 94.23%, with 97.03% specificity, and 96.39% of sensitivity. Therefore, it shows that, the improved <i>C4.5 algorithm</i> may the better ability to predict patients with CAD diseases
9	"Penerapan <i>Algoritma C4.5</i> untuk Klasifikasi Data Rekam Medis berdasarkan <i>International Classification Diseases (ICD-10)</i> ". Yudha Aditya Fiandraa , Sarjon Defit , Yuhandric. 2017 [17]	<i>Algoritma C4.5</i>	Jenis Kelamin, Kategori Umur, Bulan Berobat dan Kode Penyakit.	Hasil analisa menunjukan bahwa <i>algoritma C4.5</i> berhasil mengelompokan penyakit ke dalam 13 jenis kategori dari 21 jenis kategori yang menjadi label tujuan berdasarkan <i>ICD (International Code Diseases)</i> atau kode penyakit internasional, sehingga dapat dikatakan bahwa <i>Algoritma C4.5</i> berhasil mendefinisikan 61,9% dari kategori label tujuan yang ada.

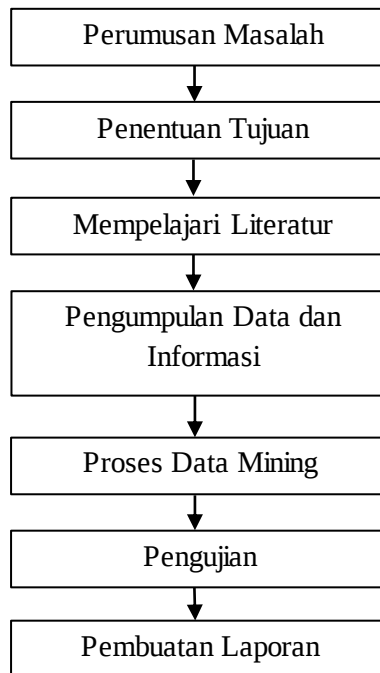
10	“Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.5.” Abdul Najib, Damar Nurcahyono, Rian Pratma Putra Setiawan. 2019 [18].	<i>Algoritma C4.5</i>	Usia, <i>GDP</i> (Gula Darah Puasa), <i>GDS</i> (Gula Darah Sewaktu) dan <i>HbA1c</i>.	Hasil prediksi error pada data yang telah diklasifikasikan bernilai akurasi 100% artinya menurut perhitungan yang dilakukan menggunakan <i>algoritma C4.5</i> tidak ada satupun error yang terjadi pada dataset, dapat disimpulkan bahwa <i>algoritma C4.5</i> memiliki akurasi yang sangat tinggi pada perhitungannya.
----	--	------------------------------	---	--

BAB III

METODOLOGI PENELITIAN

3.1. KERANGKA KERJA PENELITIAN

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya kerangka kerja (*framework*) yang jelas tahapan-tahapannya. Kerangka kerja ini merupakan langkah- langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang akan digunakan adalah sebagai berikut :



Gambar 3.1. Tahapan Penelitian

Berdasarkan kerangka kerja penelitian yang telah di gambar di atas, maka dapat di uraikan pembahasan masing masing tahapan dalam penelitian adalah sebagai berikut :

1. Perumusan Masalah

Masalah yang dirumuskan dalam penelitian ini adalah bagaimana Penerapan *Metode Decision Tree Dengan Algoritma C4.5* Untuk Klasifikasi Penyakit Jantung.

2. Penentuan Tujuan

Tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi *Algoritma C4.5* dalam klasifikasi Penyakit Jantung.

3. Mempelajari Literatur

Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, literatur-literatur bersumber dari buku-buku perpustakaan Universitas Dinamika Bangsa Jambi dan jaringan internet. Literatur-literatur yang digunakan nanti dilampirkan dalam daftar pustaka.

4. Pengumpulan Data dan Informasi

Dalam pengumpulan data, penulis mendapatkan *dataset* online yang terdapat pada sebuah website di internet.

5. Proses Data Mining

Data Mining adalah proses pengekstrasian *knowledge* yang tersimpan dalam dataset bervolume besar. Untuk mendapatkan *knowledge* dalam dataset digunakanlah *Algoritma C4.5*.

6. Pengujian

Pada tahap ini dilakukan pengujian dari hasil yang didapat dari tahap sebelumnya sebagai pedoman untuk mendapatkan hasil klasifikasi Penyakit Jantung.

7. Pembuatan Laporan

Pada tahap ini dilakukan pembuatan laporan, membuat hasil akhir dari suatu kegiatan penelitian berdasarkan data dan fakta yang telah diamati pada saat meneliti.

3.2. ANALISA TEKNIK ALGORITMA C4.5

Pada bagian ini data dan informasi yang diperoleh dan diproses dengan menggunakan Metode *Decision Tree* dengan *Algoritma C4.5* untuk mendapatkan hasil yang sesuai.

3.3. METODE PENELITIAN

Metode pengumpulan data yang digunakan dalam penelitian adalah mencari dan menggunakan *dataset* online dari *website* “www.kaggle.com” untuk dijadikan pengujian dalam penelitian ini. Pengumpulan data dilakukan untuk memperoleh hasil dan informasi yang dibutuhkan dalam hal mencapai tujuan penelitian.

3.4. ALAT PENELITIAN

Dalam membangun serta mengembangkan *website* ini, penulis menggunakan beberapa alat/piranti yang digunakan untuk melakukan pengolahan data/bahan penelitian, yaitu:

1. Hardware, dengan spesifikasi sebagai berikut:
 - a. Laptop, dengan processor AMD Radeon R7 (2.70 GHz)
 - b. RAM : 4.00 GB
 - c. HDD : 1 TB
2. Software, dengan keterangan sebagai berikut:
 - a. OS Windows 10 (64 bit)
 - b. Microsoft Excel & Word 2010
 - c. WEKA
 - d. RapidMiner Studio

3.5 JADWAL PENELITIAN

Tabel 3.1. Jadwal Penelitian

	Jenis Kegiatan	Bulan																											
		Sep				Okt				Nov				Des				Jan				Feb							
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4				
1	Perumusan Masalah																												
2	Penentuan Tujuan																												
3	Mempelajari Literatur																												
4	Pengumpulan Data & Informasi																												
5	Proses Data Mining																												
6	Pengujian																												
7	Pembuatan Laporan																												

DAFTAR PUSTAKA

- [1] R. Das, “A comparison of multiple classification methods for diagnosis of Parkinson disease,” *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1568–1572, 2010, doi: 10.1016/j.eswa.2009.06.040.
- [2] A. Rohman, V. Suhartono, and C. Supriyanto, “Penerapan Algoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung,” *J. Teknol. Inf.*, vol. 13, no. Vol 13 No 1 (2017): Jurnal Teknologi Informasi CyberKU Vol. 13, no 1, pp. 13–19, 2017.
- [3] R. Pramunendar, I. Dewi, and H. Asari, “Penentuan Prediksi Awal Penyakit Jantung Menggunakan Algoritma Back Propagation Neural Network dengan Metode Adaboost,” *Semantik*, vol. 2013, no. November, pp. 298–304, 2013, [Online]. Available: <http://publikasi.dinus.ac.id/index.php/semantik/article/view/756>
- [4] J. J. Pengaribuan, C. Tedja, and W. Sentosa, “Perbandingan Metode Algoritma C4.5 dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner,” *Informatics Eng. Res. Technol.*, vol. 1, no. 1, pp. 1–7, 2019.
- [5] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [6] E. T. L. Kusriani and E. Taufiq, “Algoritma data mining,” *Yogyakarta Andi Offset*, 2009.
- [7] J. Han and M. Kamber, “Data Mining: Concepts and Techniques, 2nd edition Morgan Kaufmann Publishers,” *San Fr. CA, USA*, 2006.
- [8] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.
- [9] F. Gorunescu, *Data Mining: Concepts, models and techniques*, vol. 12. Springer Science & Business Media, 2011.

- [10] E. Prasetyo and B. Prasetyo, "Peningkatan Akurasi Klasifikasi Algoritma C 4.5 Menggunakan Teknik Bagging pada Diagnosis Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 5, p. 1035, 2020, doi: 10.25126/jtiik.2020752379.
- [11] T. Retnasari and E. Rahmawati, "Diagnosa Prediksi Penyakit Jantung Dengan Model Algoritma Naïve Bayes Dan Algoritma C4.5," *Konf. Nas. Ilmu Sos. Teknol.*, pp. 7-12 Retnasari, T., Rahmawati, E. (2017). *Diagnos*, 2017.
- [12] E. Nurlia and U. Enri, "Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5," *J. Tek. Inform. Musirawas)* Elin Nurlia, vol. 6, no. 1, p. 42, 2021.
- [13] D. A. Firdlous, "Komparasi Algoritma Klasifikasi Data Mining untuk Memprediksi Penyakit Jantung," vol. 16, no. 1, 2022.
- [14] S. Alfaridzi, A. Nugroho, and M. R. Sani, "Penerapan data mining menggunakan algoritma C4. 5 dalam prediksi penyakit angin duduk," ... *Sist. Inf.*, vol. 13, no. 2, pp. 2383–2399, 2021, [Online]. Available: <https://ejournal.unsri.ac.id/index.php/jsi/article/view/15638>
- [15] R. Rino, "The Comparison of Data Mining Methods Using C4.5 Algorithm and Naive Bayes in Predicting Heart Disease," *Tech-E*, vol. 4, no. 2, p. 44, 2021, doi: 10.31253/te.v4i2.543.
- [16] A. A. Haruna, L. J. Muhammad, B. Z. Yahaya, E. J. Garba, N. D. Oye, and L. T. Jung, "An Improved C4.5 Data Mining Driven Algorithm for the Diagnosis of Coronary Artery Disease," *Proceeding 2019 Int. Conf. Digit. Landscaping Artif. Intell. ICD 2019*, no. 2, pp. 48–52, 2019, doi: 10.1109/ICD47981.2019.9105844.
- [17] Y. A. Fiandra, S. Defit, and Y. Yuhandri, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 2, pp. 82–89, 2017, doi: 10.29207/resti.v1i2.48.

- [18] A. Najib, D. Nurcahyono, and R. P. P. Setiawan, "Klasifikasi Diagnosa Penyakit Diabetes Mellitus (Dm) Menggunakan Algoritma C4.4," *Just TI (Jurnal Sains Terap. Teknol. Informasi)*, vol. 11, no. 2, p. 47, 2019, doi: 10.46964/justti.v11i2.153.

LAMPIRAN DATASET

Age	Sex	ChestPainType	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAngina	Oldpeak	ST_Slope	HeartDisease
40	M	ATA	140	289	0	Normal	172	N	0	Up	0
49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
37	M	ATA	130	283	0	ST	98	N	0	Up	0
48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
54	M	NAP	150	195	0	Normal	122	N	0	Up	0
39	M	NAP	120	339	0	Normal	170	N	0	Up	0
45	F	ATA	130	237	0	Normal	170	N	0	Up	0
54	M	ATA	110	208	0	Normal	142	N	0	Up	0
37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
48	F	ATA	120	284	0	Normal	120	N	0	Up	0
37	F	NAP	130	211	0	Normal	142	N	0	Up	0
58	M	ATA	136	164	0	ST	99	Y	2	Flat	1
39	M	ATA	120	204	0	Normal	145	N	0	Up	0
49	M	ASY	140	234	0	Normal	140	Y	1	Flat	1
42	F	NAP	115	211	0	ST	137	N	0	Up	0
54	F	ATA	120	273	0	Normal	150	N	1.5	Flat	0
38	M	ASY	110	196	0	Normal	166	N	0	Flat	1
43	F	ATA	120	201	0	Normal	165	N	0	Up	0
60	M	ASY	100	248	0	Normal	125	N	1	Flat	1
36	M	ATA	120	267	0	Normal	160	N	3	Flat	1
43	F	TA	100	223	0	Normal	142	N	0	Up	0
44	M	ATA	120	184	0	Normal	142	N	1	Flat	0
49	F	ATA	124	201	0	Normal	164	N	0	Up	0
44	M	ATA	150	288	0	Normal	150	Y	3	Flat	1
40	M	NAP	130	215	0	Normal	138	N	0	Up	0
36	M	NAP	130	209	0	Normal	178	N	0	Up	0
53	M	ASY	124	260	0	ST	112	Y	3	Flat	0
52	M	ATA	120	284	0	Normal	118	N	0	Up	0
53	F	ATA	113	468	0	Normal	127	N	0	Up	0
51	M	ATA	125	188	0	Normal	145	N	0	Up	0
53	M	NAP	145	518	0	Normal	130	N	0	Flat	1
56	M	NAP	130	167	0	Normal	114	N	0	Up	0
54	M	ASY	125	224	0	Normal	122	N	2	Flat	1
41	M	ASY	130	172	0	ST	130	N	2	Flat	1
43	F	ATA	150	186	0	Normal	154	N	0	Up	0
32	M	ATA	125	254	0	Normal	155	N	0	Up	0
65	M	ASY	140	306	1	Normal	87	Y	1.5	Flat	1
41	F	ATA	110	250	0	ST	142	N	0	Up	0
48	F	ATA	120	177	1	ST	148	N	0	Up	0
48	F	ASY	150	227	0	Normal	130	Y	1	Flat	0
54	F	ATA	150	230	0	Normal	130	N	0	Up	0
54	F	NAP	130	294	0	ST	100	Y	0	Flat	1
35	M	ATA	150	264	0	Normal	168	N	0	Up	0
52	M	NAP	140	259	0	ST	170	N	0	Up	0
43	M	ASY	120	175	0	Normal	120	Y	1	Flat	1
59	M	NAP	130	318	0	Normal	120	Y	1	Flat	0
37	M	ASY	120	223	0	Normal	168	N	0	Up	0
50	M	ATA	140	216	0	Normal	170	N	0	Up	0
36	M	NAP	112	340	0	Normal	184	N	1	Flat	0
....											
....											
57	M	ASY	130	131	0	Normal	115	Y	1.2	Flat	1
57	F	ATA	130	236	0	LVH	174	N	0	Flat	1
38	M	NAP	138	175	0	Normal	173	N	0	Up	0