



PROSIDING SEMINAR NASIONAL TEKNOLOGI INFORMASI & APLIKASINYA 2015

“INOVASI TEKNOLOGI INFORMASI DAN KOMUNIKASI DALAM MENUNJANG TECHNOPRENEURSHIP”

Bali, 23 Oktober 2015



Penyelenggara
PS. Teknik Informatika, Jurusan Ilmu Komputer
FMIPA - Universitas Udayana
Kampus Bukit Jimbaran, Badung Bali
Telp. (0361) 701805
<http://www.cs.unud.ac.id>

PROSIDING

**PERTEMUAN DAN PRESENTASI KARYA ILMIAH
BALI, 23 OKTOBER 2015**

**PEMBICARA UTAMA SEMINAR PANEL DENGAN TEMA
“Inovasi Teknologi Informasi dan Komunikasi dalam
Menunjang Technopreneurship”**

**Ir.Onno Widodo Purbo.M.Eng.Ph.D
Putu Sudiarta, S.Kom**

PENYUNTING AHLI

**Dr. Ahmad Ashari.M.Kom
Dr. H. Agus Zainal Arifin, S.Kom.,M.Kom
Agus Muliantara, S.Kom., M.Kom.**

PELAKSANA SEMINAR

PELINDUNG

Rektor Universitas Udayana, Bali

PENANGGUNG JAWAB

Dekan Fakultas MIPA Universitas Udayana

Ketua Program Studi Teknik Informatika, FMIPA Universitas Udayana

PANITIA

I Gusti Agung Gede Arya Kadyanan, S.Kom., M.Kom.

I Dewa Made Bayu Atmaja Darmawan, S.Kom., M.Cs.

I Wayan Supriana, S.Si., M.Cs.

Ida Bagus Made Mahendra, S.Kom., M.Kom.

I Komang Ari Mogi, S.Kom, M.Kom.

I Made Widi Wirawan, S.Si., M.Cs.

I Putu Gede Hendra Suputra, S.Kom., M.Kom.

Ngurah Agus Sanjaya ER., S.Kom., M.Kom.

Agus Muliantara, S.Kom., M.Kom.

I Made Widiartha, S.Si., M.Kom.

Made Agung Raharja, S.Si., M.Cs.

I Gusti Ngurah Anom Cahyadi Putra, S.T., M.Cs.

I Gede Santi Astawa, S.T., M.Cs.

Ida Bagus Gede Dwidasmaras, S.Kom., M.Cs.

Dra. Luh Gede Astuti, M.Kom.

DAFTAR ISI

Kata Pengantar

Daftar Isi

<i>Analisis dan Perancangan Sistem Informasi Kehadiran Pegawai pada Pusat Penelitian Perkembangan IPTEK Lembaga Ilmu Pengetahuan Indonesia</i> Warkim	1
<i>Kompresi Citra Medis dengan Wavelet Packet</i> I Made Ari Dwi Suta Atmaja.....	11
<i>Ekstraksi Fitur Warna dan Tekstur untuk Clustered-Based Retrival if Images (CLUE)</i> Sugiartha I Gusti Rai Agung.....	16
<i>Peningkatan Kemampuan Guru dalam Menggunakan Geogebra sebagai Media Pembelajaran Matematika SMP</i> Luh Putu Ida Harini	21
<i>Perbandingan Model Pembelajaran Kooperatif Tipe Belajar Bersama Menggunakan Media Google Drive dan Tanpa Google Drive</i> Desak Putu Eka Nilakusuma.....	28
<i>Analisa Faktor-Faktor yang Mempengaruhi Actual Usage dalam Penggunaan Tiket Elektronik dengan Menggunakan Technology Acceptance Model (TAM) Studi Kasus PT.KAI Commuter Jabodetabek</i> Enok Tuti Alawiah.....	35
<i>Pemanfaatan Aplikasi Google Docs sebagai Media Pembinaan Karya Ilmiah Remaja</i> Komang Dharmawan	45
<i>Penerapan WAN dengan Protokol Routing RIP dan Passive Interfaces sebagai Pemilihan Jalur Menggunakan GNS3</i> Anggarda Sanjaya	49
<i>Perancangan dan Implementasi Sistem Informasi Maajemen Rumah Sakit pada Unit Rawat Inap Studi Kasus Rumah Sakit Umum Famili Husada</i> I Dewa Ayu Kompyang Putri Utari	54
<i>Perancangan dan Implementasi Sistem Informasi Manajemen Rumah Sakit Umum Famili Husada pada Unit Poliklinik</i> I G.Ag.Sri Ag. Chandra Kusuma.....	62

<i>Sistem Temu Kembali Informasi dengan Metode TF-IDF untuk Pencarian Kata</i> I Wayan Dodik Wahyu Saputra.....	283
<i>Perancangan dan Implementasi Sistem Informasi Manajemen Laboratorium Rumah Sakit Famili Husada</i> Ida Bagus Wijana Manuaba.....	289
<i>Analisis dan Perancangan Sistem Pendukung Keputusan Pemilihan Handphone dengan Metode Entropy dan Metode Technique For Order Preference By Similarity To Ideal Solution (TOPSIS)</i> Ni Putu Sintya Dewi	298
<i>Klasifikasi Penyakit Jantung menggunakan Metode Decision Tree dengan Penerapan Algoritma C5.0</i> Sharah Islamiati	308
<i>Perancangan Algoritma Kriptografi Klasik ACK4 untuk Pengamanan Teks</i> I Nengah Tirtayasa.....	317
<i>Optimasi Lintasan Game Makepung 3D pada Engine Unity 3D</i> I Putu Agus Edy Saputra.....	322
<i>Perancangan dan Implementasi Sistem Informasi Manajemen Rumah Sakit pada Unit Rekam Medis – Studi Kasus Rumah Sakit Umu Famili Husada</i> I Putu Dharma Ade Raharja.....	328
<i>Perancangan dan Implementasi Sistem Informasi Manajemen Rumah Sakit Unit Apotek Rumah Sakit Famili Husada</i> Kadek Ary Budi Permana.....	336
<i>Sistem Pendukung Keputusan Pemilihan Guru Berprestasi (Studi Kasus : SMA N 2 Denpasar)</i> I Gusti Bagus Hadi Widhinugraha.....	343
<i>Perancangan dan Implementasi Website E-Commerce Biro Jasa Perjalanan Wisata di Bali – Studi Kasus Bali Online Trans</i> Kadek Ary Surya Wiratama.....	349
<i>Rancang Bangun Sistem Pendukung Keputusan dalam Penentuan Jurusan di Sekolah Menengah Atas (SMA) dengan Metode Simple Additive Weighting (SAW)</i> Ida Bagus Putu Trisnayana	358

KLASIFIKASI PENYAKIT JANTUNG MENGGUNAKAN METODE DECISION TREE DENGAN PENERAPAN ALGORITMA C5.0

Sharah Islamiati¹, I Made Widiartha²

Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana

Kampus Bukit Jimbaran, Badung, Bali

Email: sharah_islamiati@yahoo.com¹, madewidiartha@gmail.com²

ABSTRAK

Data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Pada penelitian ini, penulis berusaha menambang data (data mining) penyakit jantung, dimana penyakit jantung merupakan salah satu penyakit yang mematikan didunia. Dengan tujuan penelitian ini adalah untuk menguji pengaruh penggolongan penyakit jantung yang bertujuan untuk menganalisis seorang pasien menderita penyakit jantung atau tidak. Data yang ada dianalisis menggunakan pendekatan pohon keputusan (decision tree) yaitu dengan algoritma C5.0. Algoritma C5.0 adalah salah satu algoritma klasifikasi data mining yang khususnya diterapkan pada teknik decision tree. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan information gain. Pada penelitian ini pola tersebut dapat digunakan untuk menganalisis seorang pasien menderita penyakit jantung atau tidak. Didalam penelitian ini, dataset penyakit jantung di dapat dari UCI (Univercity of California Irvine) terdiri dari 270 dataset dan terdapat 13 atribut dan 2 kelas.

Kata Kunci : data mining, penyakit jantung, klasifikasi, pohon keputusan, algoritma C5.0.

ABSTRACT

Data mining is a technique that utilizes large amounts of data to obtain valuable information that was not previously known and can be used for making important decisions. In this study, the authors tried to mine the data (data mining) heart disease, in which heart disease is one of the deadliest diseases in the world. With the aim of this study was to examine the effect of heart disease classification that aims to analyze a patient suffering from heart disease or not. Existing data were analyzed using a decision tree approach (decision tree) that the C5.0 algorithm. C5.0 algorithm is one algorititma classification of data mining techniques are especially applied to the decision tree. In this algorithm selecting attributes that will be processed using information gain. In this study, the pattern can be used to analyze a patient suffering from heart disease or not. In this study, heart disease datasets obtained from the UCI (Univercity of California Irvine) consists of 270 datasets and there are 13 attributes and 2 classes.

Keywords: data mining, heart disease, classification, decision tree algorithm C5.0.

1 PENDAHULUAN

Setiap orang, dari setiap golongan, selalu mendambakan tubuh yang sehat. Permasalahan kesehatan adalah hal yang esensial bagi setiap orang, karena merupakan modal utama dalam beraktifitas sehari-hari, misal bekerja, belajar, dan bermain. Dalam kehidupan manusia, jantung adalah salah satu dari bagian terpenting yang menjaga manusia tetap hidup.

Banyak teknik yang dapat membantu dalam membangun klasifikasi prediksi penyakit jantung, pada perkembangan terbaru, Teknik- teknik yang terdapat dalam data mining mulai banyak digunakan. Khususnya teknik decision tree yang menjadi teknik yang populer karena decision tree yang dihasilkan

mudah diinterpretasikan dan divisualisasikan (Chye, 2004).

Jantung memiliki fungsi penting dalam tubuh manusia yaitu sebagai pemompa yang melakukan tekanan terhadap darah agar darah dapat mengalir ke seluruh bagian tubuh melalui pembuluh darah arteri maupun vena. Pembuluh darah arteri dan vena berguna sebagai saluran darah untuk didistribusikan oleh jantung ke seluruh tubuh dan dikembalikan ke jantung. Darah berjalan melalui sitem sirkulasi dari jantung melalui dua lengkung sirkulasi vaskuler (pembuluh darah) yang terpisah.

Penyakit jantung merupakan penyakit pembunuh nomor satu di dunia. Jumlah penderita penyakit ini terus bertambah. Menurut catatan WHO (World Health Organization) di tahun 2015, Tentunya hal tersebut dapat mengindikasikan bahwa

penyakit jantung adalah sesuatu yang tidak dapat dianggap remeh. Sayangnya, meskipun organ jantung sangat penting bagi kehidupan, jantung juga rentan terhadap penyakit. Pada zaman modern ini, dengan semakin bertambahnya polusi, gaya hidup yang serba instant, pola makan dan minum tidak sehat, sering begadang, serta berkurangnya kebutuhan manusia untuk berjalan kaki atau melakukan aktivitas fisik lainnya, mengakibatkan jantung menjadi lebih lemah. Karena hal ini, jantung manusia modern lebih rentan terhadap penyakit.

Berdasarkan teori- teori dan kenyataan di atas, maka akan mendorong diadakan penelitian “Penentuan Klasifikasi Prediksi Penyakit Jantung Menggunakan Metode Decision Tree Algoritma C5.0”. Berdasarkan dataset penyakit jantung di UCI (University of California Irvine) [9].

Dimana pada penelitian ini digunakan metode C5.0 (*Commercial Version 5.0*) untuk mengolah atribut- atribut analisis yang digunakan. C5.0 merupakan salah satu algoritma yang terdapat dalam klasifikasi data mining yang khususnya diterapkan pada teknik *Decision Tree*. Dalam algoritma C5.0 menghasilkan *tree* dengan jumlah cabang per node bervariasi. C5.0 memperlakukan variabel kontinyu dan pemilihan atribut yang akan diproses menggunakan *information gain* tertinggi sebagai *parent* bagi node selanjutnya. Dengan metode C5.0 diharapkan mampu memberikan informasi yang berguna dengan menggunakan metode tersebut, deteksi dini terhadap penyakit jantung dapat dilakukan, sehingga dapat dilakukan pencegahan atau tindakan lebih lanjut.

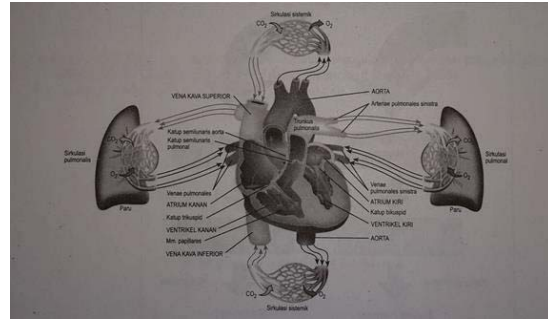
2 MODEL, ANALISIS, DESAIN, DAN IMPLEMENTASI

2.1 Landasan Teori

2.1.1 Jantung

Jantung adalah organ berongga dan memiliki empat ruang yang terletak antara kedua paru- paru di bagian tengah rongga toraks (Sloane, 2004). Jantung merupakan salah satu organ vital pada tubuh yang berfungsi untuk memompa darah keseluruh tubuh melalui pembuluh darah. Jantung terdiri dari empat ruang, yang terdiri dari empat ruang, yang terdiri dari ruang atas dinamakan atrium kanan dan kiri sedangkan ruang bawah dinamakan ventrikel kanan dan kiri.

Aliran Darah Yang Melalui Jantung



Gambar. 1 Gambaran teknis struktur jantung dan aliran darah yang melalui jantung

Sumber: Jackson, J. & Jackson, L. Clinical nursing pocket guide. Jones and Bartlett Publishers.

- Darah yang kurang mengandung oksigen (deoksigenasi) dari sirkulasi sistemik memasuki atrium kanan lalu melewati katup trikuspid menuju ventrikel kanan.
- Darah deoksigenasi ini dipompa dari ventrikel kanan melalui katup pulmonal, masuk ke arteri pulmonalis kanan dan kiri, lalu masuk ke paru.
- Di dalam paru, darah menjadi kaya oksigen lalu dialirkan oleh *venae pulmonales* menuju atrium kiri.
- Darah yang kaya oksigen dari atrium kiri masuk ke ventrikel kiri melewati katup mitral lalu melewati katup aorta, masuk ke aorta, dan dipompa ke seluruh sirkulasi sistemik.
- Aliran darah ini dipengaruhi oleh sistem konduksi listrik jantung, yang merupakan “energy listrik” sehingga jantung dapat memompa darah.

Jika jantung tidak dapat bekerja dengan baik maka dapat menyebabkan penyakit jantung.

Nyeri dada merupakan salah satu keluhan yang paling banyak ditemukan di klinik. Sebagian besar penderita merasa ketakutan bila nyeri dada tersebut disebabkan oleh penyakit jantung ataupun penyakit paru yang serius. Diagnosa yang tepat sangat tergantung dari pemeriksaan fisik yang cermat, pemeriksaan khusus lainnya serta anamnesa dari sifat nyeri dada mengenai lokasi, penyebaran, lama nyeri serta faktor pencetus yang dapat menimbulkan nyeri dada [4].

Salah satu bentuk nyeri dada yang paling sering ditemukan adalah angina pektoris yang merupakan gejala penyakit jantung koroner dan dapat bersifat progresif serta menyebabkan kematian, sehingga jenis nyeri dada ini memerlukan pemeriksaan yang lebih lanjut dan penanganan yang serius [4].

- Angina stabil (Angina klasik, Angina of Effort) : Serangan nyeri dada khas yang timbul waktu bekerja. Berlangsung hanya beberapa menit dan menghilang dengan nitroglicerine atau istirahat. Nyeri dada dapat timbul setelah makan, pada udara yang dingin, reaksi simfatis yang berlebihan atau gangguan emosi.
- Penderita dengan angina atipikal terutama wanita sering memberi hasil false positif yang tinggi. Sedangkan hasil test yang negatif pada angina atipikal dan nonangina besar kemungkinannya tidak ada kelainan koroner
- Jenis test exercise bermacam-macam antara lain test treadmill, protokol Bruce, test Master dan Sepeda ergometri. Test exercise tidak perlu dilakukan untuk diagnostik pada wanita dengan nyeri dada non anginal karena kemungkinan penyakit jantung koroner sangat rendah, sedangkan pada laki-laki dengan angina tipikal perlu dilakukan untuk menentukan penderita dengan resiko tinggi dimana sebaliknya perlu dibuat arteriografi koroner[4].

Elektrokardiogram (EKG atau ECG) member keterangan penting kepada klinis untuk menentukan kecepatan dan besarnya aktivitas listrik yang menjalar di berbagai area jantung (American Heart Association, 2007c, Twedell, 2005)[1].

2.1.2 Data Mining

Ada beberapa definisi mengenai data mining. Menurut Han (2001), Data mining merupakan sebuah proses menganalisis sekumpulan data hasil penelitian, dengan tujuan untuk menentukan hubungan antar data, dan untuk meringkas data sehingga data menjadi mudah dimengerti dan berguna bagi pemilik data [2]. Sedangkan menurut Connolly (2005: 123), data mining adalah proses menghasilkan informasi yang tidak diketahui sebelumnya dari suatu database yang besar, yang kemudian digunakan dalam proses pengambilan keputusan bisnis [5].

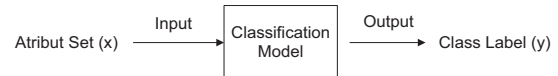
Kumpulan data dalam jumlah besar membuat organisasi memiliki data dalam jumlah banyak, namun tanpa adanya pengolahan data, data tersebut belum dapat menghasilkan informasi ataupun *knowledge* yang berguna. Oleh karena itu, data mining diperlukan untuk menemukan pola- pola dari data yang ada, meningkatkan nilai intrinsik data, dan mengubah data menjadi *knowledge* [5].

Tahapan dari proses *Knowledge Discovery in Database* (KDD) adalah :

1. Selection
2. Pre-Processing/ Cleaning
3. Transformation
4. Data mining
5. Interpretation/ Evaluation.

2.1.3 Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas- kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis.

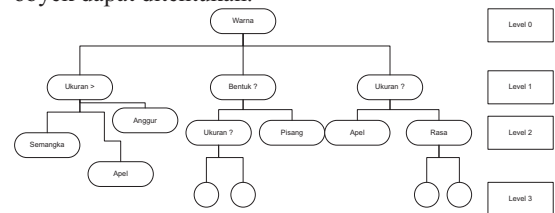


Gambar 1 : Blog Diagram Model Klasifikasi

2.1.4 Decision Tree

Decision tree adalah sebuah diagram alir yang mirip dengan struktur pohon, dimana setiap interval node notasikan atribut yang diuji, setiap cabangnya merepresentasikan hasil dari atribut tersebut, serta leaf node merepresentasikan kelas-kelas tertentu atau distribusi dari kelas-kelas [6].

Langkah ini akan berakhir pada suatu simpul jika pada simpul tersebut sudah ditemukan kelas atau jenis obyeknya. Kalau dalam satu tingkat suatu obyek telah diketahui termasuk dalam kelas tertentu, oleh karena itu berhenti di label tersebut. Jika tidak, maka dilanjutkan dengan pertanyaan di level selanjutnya hingga jelas ciri- cirinya dan jenis obyek dapat ditentukan.



Gambar . Contoh Decision Tree

Secara singkat Decision Tree merupakan salah satu metode klasifikasi pada Teks Mining. Klasifikasi adalah proses menemukan kumpulan pola atau fungsi- fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan lainnya, untuk dapat digunakan memprediksi data yang belum memiliki kelas data tertentu [6].

Pembentukan aturan keputusan, yaitu membuat aturan keputusan dari pohon yang telah dibentuk, Algoritma dasar dari decision tree diterangkan pada poin- poin berikut:

1. Susun himpunan dari kandidat partisi (S), pilih S* terbaik dari S.
2. Terangkan setiap sel (Ci) pada S*.

3. Uji kondisi terminasi pada setiap C_i yang *true* pada simpul daun.
4. *false*: ulang dengan menggunakan C_i sebagai himpunan latih yang baru.

Pada algoritma tersebut. Langkah pertama yang dilakukan adalah menentukan partisi kandidat, S, S^* dipilih dari S, S^* terbaik akan dijadikan node. Pemilihan S^* terbaik juga bergantung terhadap nilai C_i . Untuk nilai S^* lainnya yang tidak terpilih sebagai node akan dilakukan pemilihan ulang secara rekursif dengan menggunakan sisa C_i sebagai training set yang baru. Decision tree memiliki beberapa cara dalam menentukan ukuran data dalam membentuk tree, yaitu menggunakan *information gain* (untuk algoritma ID3/C4.5/C5.0).

2.1.5 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma data mining yang khususnya diterapkan pada teknik *decision tree*. C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh Ross Quinlan pada tahun 1987, yaitu ID3 dan C4.5. dalam algoritma ini pemilihan atribut yang akan diproses menggunakan *information gain*. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar. Atribut dengan nilai *information gain* tertinggi akan dipilih sebagai parent bagi node selanjutnya [10].

C5.0 adalah versi komersial dari C4.5 yang secara luas digunakan di banyak pemaketan data mining seperti Clementine and RuleQuest. Tidak seperti C4.5, penggunaan algoritma yang tepat untuk C5.0 belum terungkap. Hasil menunjukkan bahwa C5.0 meningkatkan pada penggunaan memori sekitar 90%, lebih cepat dari pada C4.5 [5].

Strategi pengembangan *decision tree* dengan menggunakan algoritma C5.0 adalah sebagai berikut:

1. Pada tahap awal, tree digambarkan sebagai node tunggal yang merepresentasikan training set.
2. Jika sampel seluruhnya berisi kelas yang sama, maka *node* tersebut menjadi *leaf* dan dilabeli dengan kelas tersebut.
3. Jika tidak, algoritma dengan menggunakan ukuran berbasis entropi (*information gain*) akan memilih variabel prediktor yang akan memisahkan record ke dalam kelas-kelas individual. Variabel tersebut menjadi variabel tes atau keputusan pada node tersebut.
4. Cabang dikembangkan untuk tiap nilai yang diketahui dari variabel tes, dan sampel dipartisi berdasarkan cabang tersebut.
5. Algoritma menggunakan proses yang sama secara rekursif membentuk *decision tree*.
6. Partisi rekursif berakhir hanya ketika satu dari kondisi-kondisi berikut terpenuhi:

- a. Seluruh record pada *node* tertentu memiliki kelas yang sama.
- b. Tidak ada atribut yang tersisa pada record yang dapat dipartisi lebih lanjut. Dalam kasus ini suara mayoritas digunakan. *Node* tersebut menjadi *leaf node* dan dilabeli dengan kelas yang menjadi mayoritas dalam record yang ada.
- c. Tidak ada record untuk cabang variabel tes. Dalam kasus ini, *leaf* terbentuk dengan mayoritas kelas sebagai label record tersebut [7].

Information Gain

Ukuran *information gain* digunakan untuk memilih atribut uji pada setiap node di dalam tree. Ukuran ini digunakan untuk memilih atribut atau node pada tree. Atribut dengan nilai *information gain* tertinggi akan terpilih sebagai parent bagi node selanjutnya. Rumus yang digunakan untuk *information gain* adalah [5]:

$$I(S_1, S_2, \dots, S_m) = -\sum_{i=1}^m p_i \cdot \log_2(p_i) \quad (1)$$

Keterangan :

S = himpunan kasus

M = jumlah sample

P_i = proporsi kelas

S_1 = jumlah sampel

S adalah sebuah himpunan yang terdiri dari s data sampel. Diketahui atribut class adalah m dimana mendefinisikan kelas-kelas di dalamnya, C_i (for $i = 1, \dots, m$), s_i adalah jumlah sampel pada S dalam class C_i . Untuk mengklasifikasikan sampel yang digunakan maka diperlukan informasi dengan menggunakan aturan seperti kelas dalam output seperti pada kelas C_i dan diestimasi dengan s_i/s .

Atribut A memiliki nilai tertentu $\{a_1, a_2, \dots, a_v\}$. Atribut A dapat digunakan pada partisi S ke dalam v subset, $\{S_1, S_2, \dots, S_v\}$, dimana S_j berisi sampel pada S yang bernilai a_j pada A . Jika A dipilih sebagai atribut tes (sebagai contoh atribut terbaik untuk split), maka subset ini akan berhubungan pada cabang dari node himpunan S . S_{ij} adalah jumlah sampel pada class C_i dalam sebuah subset S_j .

Untuk mendapatkan informasi nilai subset dari atribut A tersebut maka digunakan rumus sebagai berikut [8]:

$$E(A) = \sum_{j=1}^v \frac{S_{1j} + \dots + S_{mj}}{S} \cdot I(S_{1j}, \dots, S_{mj}) \quad (2)$$

Untuk mendapatkan nilai gain selanjutnya digunakan rumus di bawah ini [8]:

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (3)$$

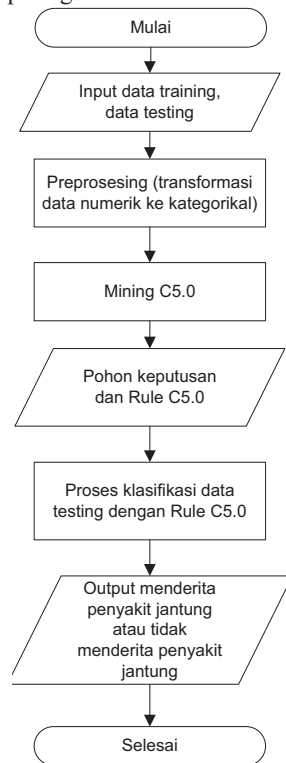
Keterangan :

- S = himpunan (dataset) kasus
A = atribut
y = jumlah partisi atribut A
 $S1j+Smj$ = jumlah kasus pada aptisi ke j (jumlah subset j)
 $I(S1, S2,...,Sm)$ = information gain
 $S1$ = jumlah sampel

2.2 Perancangan Sistem

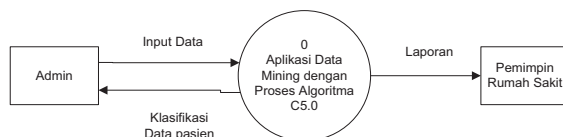
2.2.1 Flowchart Sistem

Perancangan aplikasi untuk menganalisis seorang pasien menderita penyakit jantung atau tidak. sehingga dapat dilakukan pencegahan atau tindakan lebih lanjut dengan perancangan conceptual flowchart seperti gambar berikut ini.



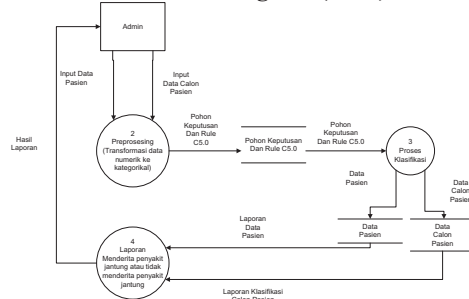
Gambar. Flowchart Sistem

2.2.2 Diagram Konteks



Gambar. Diagram Konteks

2.2.3 Data Flow Diagram (DFD) Level 0



Gambar. Data Flow Diagram (DFD) Level 0

3 DATA PENELITIAN

Dataset yang digunakan pada percobaan ini adalah data penyakit jantung yang di dapat dari UCI (University of California Irvine)[9]. Terdapat dari 270 dataset terdapat 13 atribut dan 2 kelas yaitu :

- Usia : Usia dalam Tahun
- Jenis Kelamin : Seks (1 = Laki- laki, 0 = Perempuan)
- Jenis sakit di dada/ jenis nyeri dada
 Nilai 1 : Angina yang khas
 Nilai 2 : Angina Atipikal
 Nilai 3 : Nyeri non – angina
 Nilai 4 : Asimtomatik
- Tekanan darah saat beristirahat (mm Hg pada masuk ke rumah sakit)
- Cholestoral serum dalam mg / dl
- Gula darah saat istirahat > 120 mg / dl (1 = true, 0 = false)
- Hasil elektrokardiografi saat istirahat (kelainan pada aktifitas elektrik jantung melalui gelombang irama jantung)
 Nilai 0 : Normal
 Nilai 1 : Memiliki gelombang kelainan ST – T (T Inversi gelombang dan / atau elevasi ST atau depresi dari > 0,05 mv)
 Nilai 2 : Menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri dengan criteria estes
- Detak jantung maksimum dicapai
- Latihan diinduksi angina (jenis nyeri di dada yang disebabkan oleh berkurangnya aliran darah ke otot jantung) (1 = ya, 0 = tidak ada)
- Oldpeak = depresi ST yang diakibatkan oleh latihan relative terhadap saat istirahat.
- Kemiringan segmen latihan puncak ST
 Nilai 1 : Condong keatas
 Nilai 2 : Datar
 Nilai 3 : Sedikit landai

12. Banyaknya nadi utama (0-3) yang diwarnai oleh flourosopy

13. Kondisi : 3 = Normal, 6 = Cacat Tetap, 7 = Cacat sementara.

Dan 2 tipe kelas yaitu :

Nilai 1 : Untuk tidak menderita penyakit jantung

Nilai 2 : Untuk menderita penyakit jantung

Untuk melakukan eksperimen, dari seluruh dataset yang dimiliki, akan digunakan 70%-nya sebagai *data training* dan 30%-nya sebagai *data testing*.

4 HASIL UJI COBA

Tahap awal cara kerja dari proses perhitungan Algoritma C5.0 adalah dengan melakukan pengambilan data training dari data pasien. Adapun atribut-atribut penentu yang digunakan dalam mengklasifikasikan data pasien yaitu :

1. Usia.
Merupakan atribut usia pasien yang dikelompokkan dalam sebelas kategori yaitu 23.7 – 29 Tahun, 29 – 34.3 Tahun, 34.3 – 39.6 Tahun, 39.6 – 44.9 Tahun, 44.9 – 50.2 Tahun, 50.2 – 55.4 Tahun, 55.4 – 60.7 Tahun, 60.7 – 66 Tahun, 66 – 71.3 Tahun, 71.3 – 76.6 Tahun, dan 76.6 – 81.9 Tahun
2. Jenis Kelamin.
Merupakan atribut jenis kelamin pasien yang dikelompokkan dalam dua kategori yaitu 0.0 Perempuan dan 1.0 laki – laki.
3. Jenis Sakit didada
Merupakan atribut jenis sakit didada yang dikelompokkan dalam empat kategori yaitu 1.0 Angina yang khas, 2.0 Angina Atipikal, 3.0 Nyeri non – angina, dan 4.0 Asimtomatik.
4. Tekanan darah saat beristirahat
Merupakan atribut tekanan darah saat beristirahat yang dikelompokkan dalam sebelas kategori yaitu 82.3 – 94, 94 – 105.7, 105.7 – 117.4, 117.4 – 129, 129 – 140.7, 140.7 – 152.4, 152.4 – 164.1, 164.1 – 175.7, 175.7 – 187.4, 187.4 – 199.1, dan 199.1 – 210.8
5. Cholesterol serum dalam mg / dl
Merupakan atribut kolesterol serum yang dikelompokkan dalam sembilan kategori yaitu 77.7 – 126, 126 – 174.3, 174.3 – 222.5, 222.5 – 270.8, 270.8 – 319, 319 – 367.3, 367.3 – 415.5, 415.5 – 463.8, dan 463.8 – 608.5
6. Gula darah saat istirahat
Merupakan atribut gula darah saat istirahat yang dikelompokkan dalam dua kategori yaitu 0.0 false dan 1.0 true.
7. Hasil elektrokardiografi saat istirahat
Merupakan atribut hasil elektrokardiografi yang dikelompokkan dalam tiga kategori yaitu 0.0 Normal, 1.0 Memiliki gelombang kelainan

ST – T, dan 2.0 menunjukkan kemungkinan atau pasti hipertrofi ventrikel kiri dengan kriteria estes.

8. Detak jantung maksimum dicapai
Merupakan atribut detak jantung maksimum yang dikelompokkan dalam sebelas kategori yaitu 56.6 – 71, 71- 85.4, 85.4 – 99.9, 99.9 – 114.3, 114.3 – 128.7, 128.7 – 143.2, 143.2 – 157.6, 157.6 – 172, 172 – 186.5, 186.5 – 200.9, dan 200.9 – 215.3
9. Latihan diinduksi angina
Merupakan atribut latihan diinduksi angina yang dikelompokkan dalam dua kategori yaitu 0.0 tidak ada dan 1.0 ya.
10. Oldpeak
Merupakan atribut oldpeak yang dikelompokkan dalam sembilan kategori yaitu 0 – 0.7, 0.7 – 1.4, 1.4 – 2, 2 – 2.7, 2.7 – 3.4, 3.4 – 4.1, 4.1 – 4.8, 4.8 – 6.1, dan 6.1 – 6.8
11. Kemiringan segmen latihan puncak ST
Merupakan atribut kemiringan segmen latihan puncak ST yang dikelompokkan dalam tiga kategori yaitu 1.0 condong keatas, 2.0 datar, dan 3.0 sedikit landai
12. Banyak nadi utama yang diwarnai oleh flourosopy
Merupakan atribut banyaknya nadi utama yang dikelompokkan dalam empat kategori yaitu 0.0, 0.1, 0.2, dan 0.3
13. Kondisi
Merupakan atribut kondisi yang dikelompokkan dalam tiga kategori yaitu 3.0 normal, 6.0 cacat tetap, dan 7.0 cacat sementara

Setiap atribut memiliki nilai. Sedangkan kelasnya ada pada kolom Keterangan yaitu kelas “Menderita Penyakit Jantung” dan kelas “Tidak Menderita Penyakit Jantung”. Kemudian data tersebut dianalisis; dataset tersebut memiliki 270 kasus yang terdiri 120 “Menderita Penyakit Jantung” dan 150 “Tidak Menderita Penyakit Jantung”. Untuk melakukan eksperimen, dari seluruh dataset yang dimiliki, akan digunakan 70%-nya sebagai *data training* yaitu 189 kasus.

Kemudian hitung information gain dengan rumus sebagai berikut :

$$I(S_1, S_2, \dots, S_m) = -\sum_{i=1}^m p_i * \log_2(p_i) \quad (1)$$

Keterangan :

S = himpunan kasus

M = jumlah sample

Pi = proporsi kelas

S1= jumlah sampel

Sharah Islamiati, Klasifikasi Penyakit Jantung Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C5.0

Jadi

$$I(S_1, S_2, \dots, S_m) = \left(- \left(\frac{85}{189} \right) \log_2 \left(\frac{85}{189} \right) \right) + \left(- \left(\frac{104}{189} \right) \log_2 \left(\frac{104}{189} \right) \right) = 0.992698$$

Tabel 2. Hasil Perhitungan pada Dataset

Total Kasus	Kelas (Menderita Penyakit Jantung)	Kelas (Tidak Menderita Penyakit Jantung)	$I(S_1, S_2, \dots, S_m)$ Total Informasi Gain
189	85	104	0,992698

Setelah mendapatkan informasi gain dari keseluruhan kasus, lakukan analisi pada setiap atribut dan nilai – nilainya dan hitung informasi gainnya seperti yang ditampilkan pada Tabel di bawah ini.

Tabel Analisis Atribut, Nilai, Banyaknya Kejadian Nilai, Informasi Gain dan Gain

Node	Atribut	Nilai	Sum (Nilai)	Sum (Ya)	Sum (Tidak)	Informasi Gain	Gain
1	Usia	26.4 - 31.6	0	0	0	0	
		31.6 - 36.9	3	1	2	0.918295834	
		36.9 - 42.2	21	5	16	0.791858353	
		42.2 - 47.5	25	7	18	0.855450811	
		47.5 - 52.8	29	11	18	0.957553484	
		52.8 - 58.1	44	26	18	0.976020648	
		58.1 - 63.4	33	22	11	0.918295834	
		63.4 - 68.7	26	11	15	0.98285869	
		68.7 - 73.9	6	2	4	0.918295834	
		73.9 - 79.2	2	0	2	0	
	Jenis Kelamin	0	65	15	50	0.779349837	
		1	124	70	54	0.987956501	
	Jenis Sakit Jidada	1	15	4	11	0.836640742	
		2	28	5	23	0.67694187	
		3	51	11	40	0.752212173	
		4	95	65	30	0.899743759	
							0.821918
	Tekanan Darah	88.2 - 99.8	1	0	1	0	
		99.8 - 111.5	26	10	16	0.961236605	
		111.5 - 123.2	44	17	27	0.962412735	
		123.2 - 134.9	47	20	27	0.983939395	
		134.9 - 146.6	38	18	20	0.998000884	
		146.6 - 158.2	18	11	7	0.964078765	
		158.2 - 169.9	7	3	4	0.985228136	
		169.9 - 181.6	6	4	2	0.918295834	
		181.6 - 193.3	1	1	0	0	
		193.3 - 204.9	1	1	0	0	
							0.959086
	Cholestrol	101.9 - 150.1	4	1	3	0.811278124	
		150.1 -	22	8	14	0.945660305	

	198.4 - 246.6	82	31	51	0.956652272	
	246.6 - 294.9	50	31	19	0.958042022	
	294.9 - 343.1	24	12	12	1	
	343.1 - 391.4	2	0	2	0	
	391.4 - 439.7	4	2	2	1	
	439.7 - 536.2	1	0	1	0	
	536.2 - 584.4					0.926731
Gula Darah	0	159	71	88	0.991738126	
	1	30	14	16	0.996791632	
						0.99254
Elektrokardiografi	0	90	33	57	0.948078244	
	1	2	1	1	1	
	2	97	51	46	0.998082508	
						0.974291

Detak Jantung	63.8 - 78.2	1	1	0	0	
	78.2 - 92.6	1	1	0	0	
	92.6 - 107.1	10	8	2	0.721928095	
	107.1 - 121.5	18	13	5	0.852405179	
	121.5 - 135.9	24	16	8	0.918295834	
	135.9 - 150.4	36	19	17	0.997772472	
	150.4 - 164.8	48	14	34	0.870864469	
	164.8 - 179.2	39	10	29	0.821280942	
	179.2 - 193.7	10	2	8	0.721928095	
	193.7 - 208.1	2	1	1	0	
						0.854879
Angina	0	124	40	84	0.907165768	
	1	65	45	20	0.89049164	
						0.901431
Oldpeak	- 0.3 - 0.3	25	10	15	0.970950594	
	0.3 - 1	17	4	13	0.787126586	

	1 - 1.7	19	11	8	0.981940787	
	1.7 - 2.4	6	2	4	0.918295834	
	2.4 - 3.1	11	5	6	0.994030211	
	3.1 - 3.8	2	1	1	1	
	3.8 - 4.4	0	0	0	0	
	5.1 - 5.8	0	0	0	0	
	5.8 - 6.5	1	1	0	0	
						0.395534
Kemiringan	1	90	22	68	0.802353443	
	2	86	57	29	0.922123131	
	3	13	6	7	0.995727452	
						0.870153
Banyaknya Nadi	0	108	26	82	0.796269937	
	1	43	27	16	0.952265625	
	2	25	21	4	0.634309555	
	3	13	11	2	0.619382195	
						0.798171
Kondisi	3	111	26	85	0.78531981	
	6	10	7	3	0.881290899	
	7	68	52	16	0.787126586	
						0.791048

Untuk mendapatkan informasi nilai subset dari atribut A tersebut maka digunakan rumus sebagai berikut [8]:

$$E(A) = \sum_{j=1}^s \frac{S_{1j} + \dots + S_{mj}}{s} | (S_{1j}, \dots, S_{mj}) \quad (2)$$

$$S_{1j} + S_{mj}$$

Keterangan : S = jumlah subset j yang dibagi dengan jumlah sampel S (jumlah kasus dalam S).

Jadi :

$$E(Umur) = \left(\frac{0}{189} \times 0 + \frac{3}{189} \times 0.918295834 + \frac{21}{189} \times 0.791858353 + \frac{25}{189} \times 0.855450811 + \frac{29}{189} \times 0.957553484 + \frac{44}{189} \times 0.976020648 + \frac{33}{189} \times 0.918295834 + \frac{26}{189} \times 0.98285869 + \frac{6}{189} \times 0.918295834 + \frac{2}{189} \times 0 \right) = 0.078137$$

Untuk mendapatkan nilai gain selanjutnya digunakan rumus di bawah ini [8]:

$$Gain(A) = I(S_1, S_2, \dots, S_m) - E(A) \quad (3)$$

Keterangan :

S = himpunan (dataset) kasus

A = atribut

y = jumlah partisi atribut A

$S_{1j} + S_{mj}$ = jumlah kasus pada partisi ke j (jumlah subset j)

S_1 = jumlah sampel

Jadi :

$$Gain(Umur) = 0.992698 - 0.078137 = 0.914561$$

Hitung pula Gain (Jenis Kelamin), Gain (Jenis Sakit di Dada), Gain (Tekanan Darah Saat Beristirahat), Gain (Cholesterol Serum), Gain (Gula Darah Saat Beristirahat), Gain (Hasil Elektrokardiografi), Gain (Detak Jantung Maksimum Dicapai), Gain (Latihan Diinduksi Angina), Gain (Oldpeak), Gain (Kemiringan Segmen), Gain (Banyak Nadi Utama), Gain (Kondisi). Nilai gain terbesar menjadi node akar (root node) menjadi daun atau leaf.

Setelah pembentukan pohon keputusan dianalisis lebih lanjut. Untuk mempermudah difilter dengan mengambil data.

Kemudian data di dianalisis dan dihitung lagi informasi gain dan informasi gain setiap atribut serta gainnya. Serta itu tentukan pilihan atribut yang memiliki gain tertinggi untuk dibuatkan node berikutnya.

Salah satu dari kondisi di bawah ini terpenuhi Partisi rekursif berakhir

- Seluruh record pada *node* tertentu memiliki kelas yang sama.
- Tidak ada atribut yang tersisa pada record yang dapat dipartisi lebih lanjut. Dalam kasus ini suara mayoritas digunakan. *Node*

tersebut menjadi *leaf node* dan dilabeli dengan kelas yang menjadi mayoritas dalam record yang ada.

- Tidak ada record untuk cabang variabel tes. Dalam kasus ini, *leaf* terbentuk dengan mayoritas kelas sebagai label record tersebut

Dari hasil eksperimen di atas dapat dilihat nilai akurasi yaitu 90,08%

5 KESIMPULAN

Berdasarkan hasil pembahasan maka penulis dapat mengambil beberapa kesimpulan dari penelitian dan pengembangan antara lain :

- Sistem klasifikasi data pasien ini digunakan untuk menampilkan informasi klasifikasi menderita penyakit jantung atau tidak menderita penyakit jantung dengan menggunakan metode Decision Tree dengan penerapan algoritma C5.0.
- Atribut penentu yang digunakan dalam penelitian ini terdiri dari 13 atribut yaitu : Usia,, Jenis Kelamin, Jenis sakit di dada/ jenis nyeri dada, Tekanan darah saat beristirahat (mm Hg pada masuk ke rumah sakit), Cholesterol serum dalam mg / dl, Gula darah saat istirahat > 120 mg / dl (1 = true, 0 = false), Hasil elektrokardiografi saat istirahat (kelainan pada aktifitas elektrik jantung melalui gelombang irama jantung), Detak jantung maksimum dicapai, Latihan diinduksi angina (jenis nyeri di dada yang disebabkan oleh berkurangnya aliran darah ke otot jantung), Oldpeak = depresi ST yang diakibatkan oleh latihan relative terhadap saat istirahat., Kemiringan segmen latihan puncak ST, Banyaknya nadi utama (0-3) yang diwarnai oleh flourosopy, Kondisi : 3 = Normal, 6 = Cacat Tetap, 7 = Cacat sementara.
- Kelas- kelas penentu yang di gunakan dalam penelitian ini terdiri dari 2 kelas yaitu Nilai 1 : Untuk tidak menderita penyakit jantung dan Nilai 2 : Untuk menderita penyakit jantung
- Dimana dengan adanya sistem ini maka mempermudah pihak Rumah Sakit dalam memperkirakan pasien menderita penyakit jantung atau tidak. sehingga dapat dilakukan pencegahan atau tindakan lebih lanjut.
- Sistem ini dapat dikembangkan dengan menambahkan metode pembandingan untuk mendapatkan hasil akurasi yang lebih baik.
- Tampilan dari aplikasi ini masih terlihat kaku, sehingga dapat dibuat lebih menarik lagi.

6. DAFTAR PUSTAKA

- [1] Theresa Ann Middleton Brosche. 2014. **Buku Saku EKG**. Jakarta : Penerbit Buku Kedokteran EGC.
- [2] Han, Jiawei dan Khamber, Micheline. 2001. **Data Mining : Concepts and Techniques**. Morgan Kaufmann Publishers. San Francisco.
- [3] Connolly, Thomas dan Begg, Carolyn E. (2005). **Database System: A Practical Approach to Design, Implementation, and Management**. Fourth Edition. Addison Wesley Publishing company, Inc, USA.
- [4] T.Bahri Anwar. 2004. "Nyeri Dada". Fakultas Kedokteran Universitas Sumatra Utara.
- [5] Kurniarti Putri Wirdhaningsih., Dian Eka Ratnawati, S.Si., M.Kom., Drs. Marji, MT. "Penerapan Algoritma Decision Tree C5.0 Untuk Peramalan Forex". Universitas Brawijaya Malang.
- [6] Pramudya Prima Insan., Achmad Ridok., dan Indriati. "Klasifikasi Emosi Untuk Teks Berbahasa Indonesia Dengan Menggunakan Algoritma C5.0". Universitas Brawijaya Malang
- [7] Yogi Yusuf W., "Perbandingan Performansi Algoritma Decision Tree C5.0, CART, Dan AHAIID: Kasus Prediksi Status Resiko Kredit Di Bank X". SNATI 2007.
- [8] Prof. Nilima Patil., Prof. Rekha Lathi., Prof. Vidya Chitre. 2012. "Comparison of C5.0 & CART Classification algorithms using pruning technique. International Journal of Engineering Research & Technology (IJERT)". Vol. 1. ISSN: 2278-0181. [4 June 2012].
- [9] **Universitas of California Irvine Machine Learning Repository**,
<URL:<http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/heart.dat>>.
- [10] Ernawati, Iin., 2008. **Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 Dan K-Nearest Neighbor**.
<URL:<http://dokumen.tips/documents/2008ier-abstract.html>>.



Penyelenggara
PS. Teknik Informatika, Jurusan Ilmu Komputer
FMIPA - Universitas Udayana
Kampus Bukit Jimbaran, Badung Bali
Telp. (0361) 701805
<http://www.cs.unud.ac.id>