

---

## IMPLEMENTASI ALGORITMA DECISION TREE C4.5 DALAM PERANCANGAN SISTEM INFORMASI DATA REKAM MEDIS PENYAKIT JANTUNG

Andi

STMIK TIME Medan

Jl. Merbabu No.32 AA-BB Medan 20212, Telp:061-4561932

E-mail: andi@stmik-time

---

**Abstrak** – Jantung merupakan otot yang terbagi menjadi empat ruang berupa atrium (serambi) kanan dan kiri di bagian atas sedangkan dua ruang lagi terletak di bagian bawah, yaitu ventrikel (bilik) kanan dan kiri. Pada praktiknya, di dunia pendidikan kedokteran seringkali data rekam medis penyakit jantung disimpan dengan tujuan untuk proses pembelajaran ataupun diolah menjadi sebuah pengetahuan. Pada website <https://www.kaggle.com/>, banyak disediakan dataset yang disimpan dan diolah dengan tujuan pembelajaran, meliputi dataset penyakit, dataset populasi ataupun jenis dataset lainnya. Namun kenyataannya, untuk mengolah data rekam medis tersebut menjadi sebuah pengetahuan tidaklah mudah, dikarenakan banyaknya jumlah data rekam medis yang tercatat tidak memungkinkan bagi manusia dalam mengolahnya. Selain itu, pengelolaan data rekam medis secara konvensional memiliki tingkat keakuratan yang kurang baik sehingga hasil kesimpulan yang diambil pun tentunya akan berbeda dan tidak begitu akurat. Oleh karena permasalahan tersebut, maka diperlukan penerapan teknik data mining dalam pada bidang medis, khususnya pengelolaan data rekam medis. Pada penelitian ini, algoritma data mining yang digunakan adalah Decision Tree C4.5. Algoritma ini dipilih dikarenakan cukup akurat dan kompleks dalam melakukan proses pengelolaan data rekam medis. Hasil penelitian menunjukkan bahwa penerapan algoritma Decision Tree C4.5 tergolong cukup akurat dalam melakukan prediksi terhadap data rekam medis penyakit jantung dimana hasil pengujian dengan mendapatkan akurasi sebesar 96%.

**Kata Kunci** : *Sistem Informasi Klarifikasi Data Rekam Medis, Penyakit Jantung, Data Mining, Algoritma Decision Tree C4.5*

---

### 1. LATAR BELAKANG

Jantung merupakan otot yang terbagi menjadi empat ruang berupa *atrium* (serambi) kanan dan kiri di bagian atas sedangkan dua ruang lagi terletak di bagian bawah, yaitu *ventrikel* (bilik) kanan dan kiri. Antara ruang kanan dan kiri dipisahkan oleh *septum* (dinding otot) yang berfungsi mencegah tercampurnya darah yang kaya oksigen dengan darah yang miskin oksigen. Jantung sangat rentang mengalami gangguan serta mengalami penyakit. Penyakit jantung merupakan kondisi ketika jantung mengalami gangguan. Bentuk gangguan pada jantung bisa bermacam-macam yaitu seperti adanya gangguan pada pembuluh darah jantung, irama jantung, katup jantung, atau gangguan akibat bawaan lahir. Pada praktiknya, di dunia kedokteran seringkali data rekam medis penyakit jantung disimpan dengan tujuan proses pembelajaran ataupun diolah menjadi sebuah pengetahuan. Pada website <https://www.kaggle.com/>, banyak disediakan *dataset* yang disimpan dan diolah dengan tujuan pembelajaran, baik itu *dataset* penyakit, *dataset* populasi ataupun jenis *dataset* lainnya. Namun kenyataannya, untuk mengolah data rekam medis tersebut menjadi sebuah pengetahuan tidaklah mudah, dikarenakan banyaknya jumlah data rekam medis yang tercatat tidak memungkinkan bagi manusia dalam mengolahnya. Selain itu, pengelolaan data rekam medis secara konvensional memiliki tingkat keakuratan yang kurang baik sehingga hasil kesimpulan yang diambil pun tentunya akan berbeda dan tidak begitu akurat. Oleh karena permasalahan tersebut, maka diperlukan penerapan teknik *data mining* dalam dunia kesehatan khususnya pengelolaan data rekam medis. *Data mining* merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. *data mining* itu sendiri disalurkan dalam bentuk algoritma-algoritma kompleks yang memiliki tingkat keakuratan yang cukup baik. Terdapat beberapa penelitian yang mencoba membangun sistem pengelolaan klasifikasi data rekam medis yaitu penelitian yang dilakukan oleh Edy Kurniawan pada tahun 2017. Pada penelitian tersebut dilakukan analisa data rekam medis menggunakan teknik *Data Mining Association* dengan algoritma *Clustering* [1]. Selanjutnya penelitian yang dilakukan oleh Amir Ali pada tahun 2019 yaitu membangun sebuah sistem klusterisasi data rekam medis dengan menggunakan algoritma *K-Means Clustering* [2]. Kedua penelitian tersebut sudah cukup baik, namun masih memiliki kekurangan pada tingkat akurasi. Pada penelitian ini, algoritma *data mining* yang digunakan adalah *Decision Tree C4.5*. Algoritma ini dipilih dikarenakan cukup akurat dan kompleks dalam melakukan proses pengelolaan data rekam medis [3].

2. LANDASAN TEORI

**Klasifikasi**

Klasifikasi merupakan suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Aturan-aturan tersebut digunakan pada data baru untuk diklasifikasi. Teknik ini menggunakan *supervised induction*, yang memanfaatkan kumpulan pengujian *record* yang terklasifikasi untuk menentukan kelas-kelas tambahan [4].

**Data Rekam Medis**

Rekam medis adalah keterangan, data dan informasi mengenai demografis pasien, sejarah medis pasien serta tindakan tindakan medis yang pernah diberikan kepada pasien bersangkutan. Rekam medis bukan hanya merupakan suatu pencatatan, tetapi suatu sistem yang berjalan meliputi proses perekaman, penyimpanan serta penyajian rekam medis tersebut kepada pihak-pihak yang membutuhkan [5].

**Data Mining**

Data *mining* bisa diartikan proses pencarian pola data yang tidak diketahui atau tidak diperkirakan sebelumnya. Data *mining* merupakan salah satu tahapan dalam keseluruhan proses *knowledge discovery in database* [6].

**Algoritma Decision Tree C4.5**

*Decision tree* atau pohon keputusan merupakan salah satu teknik klasifikasi. *Decision tree* adalah *top-down* pohon rekursif dari algoritma induksi, yang menggunakan ukuran seleksi atribut untuk memilih atribut yang diuji. Dengan pohon keputusan, manusia dapat dengan mudah melihat hubungan antara faktor-faktor yang mempengaruhi suatu masalah. Pohon keputusan ini juga dapat menganalisa nilai suatu informasi yang terdapat dalam suatu alternatif pemecahan masalah. Pohon keputusan adalah salah satu metode klasifikasi yang mudah untuk diaplikasikan oleh manusia. Pohon keputusan bisa melakukan prediksi dengan menggunakan struktur pohon. Pohon keputusan mengubah data menjadi pohon keputusan serta menjadi beberapa aturan keputusan. Gambar 2.2 menunjukkan struktur pohon keputusan [7].

Terdapat formula ukuran atribut C4.5 antara lain: [6]

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \tag{1}$$

Dengan Info (D) merupakan informasi yang dibutuhkan untuk mengklasifikasikan label kelas sebuah *tuple* di D.  $p_i$  adalah peluang bukan nol dengan sebuah *tuple* acak di D. Fungsi log menggunakan basis 2, karena informasi yang dikodekan dalam bit. Info (D) juga dikenal sebagai *entropy*. Nilai *entropy* yang dihasilkan untuk mengklasifikasi *tuple* dari D berdasarkan partisi oleh A adalah:

$$Info_A(D) = \sum_{J=1}^Y \frac{|D_j|}{D} x Info(D_j) \tag{2}$$

Untuk mendapatkan nilai *Information Gain* pada atribut A selanjutnya digunakan rumus:

$$GAIN(A) = Info(D) - Info(D_j) \tag{3}$$

Gain (A) menyatakan berapa banyak cabang yang akan diperoleh pada A. Atribut A dengan *information gain* tertinggi. *Information Gain* (A), dipilih sebagai atribut pada *node* N.

**Confusion Matrix**

*Confusion Matrix* merupakan dataset yang memiliki dua kelas, kelas yang satu sebagai positif dan kelas yang lain sebagai negatif. Terdiri dari empat sel yaitu *True Positives* (TP), *False Positives* (FP), *True Negatives* (TN), dan *False Negatives* (FN) [8].

|              |            | Predicted Class     |                     |
|--------------|------------|---------------------|---------------------|
|              |            | Faulty              | Not Faulty          |
| Actual Class | Faulty     | True Positive (TP)  | False Negative (FN) |
|              | Not Faulty | False Positive (FP) | True Negative (TN)  |

Gambar 1. *Confusion Matrix* Untuk Dua Model Kelas [8]

Untuk menghitung akurasi menggunakan rumus: [8]

$$Accuracy = \frac{a+d}{a+b+c+d} = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

### 3. METODE PENELITIAN

#### Metode Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini adalah metode studi pustaka yaitu mengumpulkan data teori melalui jurnal, media cetak, ataupun sumber-sumber referensi dari internet. Penelitian ini menggunakan *dataset* penyakit jantung yang berasal dari <https://www.kaggle.com/ronitf/heart-disease-uci> dengan jumlah dataset sebanyak 304 data.

#### Analisis Sistem

Analisis sistem pada penelitian ini terbagi menjadi 3 tahapan proses, yaitu:

- a. Analisis sistem berjalan yaitu melakukan analisis sistem berjalan yang digunakan pada saat ini, khususnya proses klasifikasi data rekam medis penyakit jantung.

Berikut ini akan diuraikan analisis sistem berjalan pada penelitian ini yaitu:

- 1) Pada praktiknya, proses pengelolaan data rekam medis masih dilakukan secara konvensional oleh manusia sehingga seringkali terjadinya kesalahan klasifikasi terhadap data rekam medis tersebut.
- 2) Selain secara konvensional, proses pengelolaan data rekam medis memanfaatkan aplikasi, namun aplikasi yang tersedia masih belum memanfaatkan algoritma dalam proses pengelompokan sehingga hasil klasifikasi cenderung tidak akurat.

Berdasarkan hasil analisis sistem berjalan tersebut, maka perlu diusulkan sebuah solusi yang dapat menyelesaikan kekurangan pada proses berjalan saat ini. Solusinya adalah dengan membangun sebuah sistem informasi data rekam medis yang menerapkan algoritma *Data Mining Decision Tree C4.5* dalam melakukan pengklasifikasian. Melalui algoritma C4.5, maka sistem akan secara otomatis mempelajari data rekam medis baru yang akan masuk dan kemudian mengklasifikasikannya apakah data tersebut merupakan pasien penyakit jantung atau pasien yang sehat.

- b. Analisis algoritma yang digunakan yaitu algoritma *Decision Tree C4.5* dengan mengambil sampel berupa 10 data *training* kemudian 1 data pasien yang akan diklasifikasikan melalui penerapan algoritma tersebut. Algoritma C4.5 akan mempelajari data pasien penyakit jantung, kemudian melalui hasil pembelajaran tersebut maka akan diolah dan ditentukan apakah data rekam medis pasien tersebut merupakan *Healthy* (0) atau *Sick* (1).

Sebagai contoh kasus sederhana, akan disediakan 10 data rekam medis pasien sebagai data *training* dan peserta ke 11 akan dilakukan klasifikasi.

Tabel 1. Contoh Data *Training* Pasien Penyakit Jantung Beserta Kriterianya

| N  | Ag | Se | C | Trestbp | Cho | FB | Restec | Thalac | Exan | Slop | Target/Klasifik | Nama    |
|----|----|----|---|---------|-----|----|--------|--------|------|------|-----------------|---------|
| o  | e  | x  | P | s       | l   | S  | g      | h      | g    | e    | asi             | Pasien  |
| 1  | 63 | 1  | 3 | 145     | 233 | 1  | 0      | 150    | 0    | 0    | 1               | Andi    |
| 2  | 37 | 1  | 2 | 130     | 250 | 0  | 1      | 187    | 0    | 0    | 1               | Budi    |
| 3  | 41 | 0  | 1 | 130     | 204 | 0  | 0      | 172    | 0    | 2    | 1               | Cindy   |
| 4  | 67 | 1  | 0 | 160     | 286 | 0  | 0      | 108    | 1    | 1    | 0               | Daddi   |
| 5  | 67 | 1  | 0 | 120     | 229 | 0  | 0      | 129    | 1    | 1    | 0               | Erdi    |
| 6  | 62 | 0  | 0 | 140     | 268 | 0  | 0      | 160    | 0    | 0    | 0               | Fredick |
| 7  | 56 | 1  | 1 | 120     | 236 | 0  | 1      | 178    | 0    | 2    | 1               | Gigi    |
| 8  | 57 | 0  | 0 | 120     | 354 | 0  | 1      | 163    | 1    | 2    | 1               | Herry   |
| 9  | 57 | 1  | 0 | 140     | 192 | 0  | 1      | 148    | 0    | 1    | 1               | Iriani  |
| 10 | 56 | 0  | 1 | 140     | 294 | 0  | 0      | 153    | 0    | 1    | 1               | Jimmy   |
| 11 | 64 | 1  | 3 | 110     | 211 | 0  | 0      | 144    | 1    | 1    | ?               | Nindy   |

Sumber: <https://www.kaggle.com/ronitf/heart-disease-uci>

Keterangan:

- *Age*: Usia dalam tahun (numerik).
- *Sex*: Jenis kelamin dengan nilai 1: Laki-laki, 0: Perempuan.
- *Chest pain type* (CP) dengan nilai 0: Normal, 1: Typical Angina, 2: Atypical Angina, 3: Non-Anginal Pain, 4: Asymptomatic.
- *Trestbps*: *resting blood pressure* pasien dalam mm Hg pada saat masuk rumah sakit (numerik).
- *Chol*: Serum kolesterol dalam mg/dl (numerik).
- *Fbs*: Ukuran *boolean* yang menunjukkan apakah *fasting blood sugar* lebih besar dari 120 mg/dl atau tidak dengan nilai 1: *True*; 0: *False*.
- *Restecg*: Hasil ECG selama istirahat dengan nilai 0: normal, 1: abnormal (memiliki kelainan gelombang ST-T), 2: hipertrofi ventrikel.
- *Thalac*: detak jantung maksimum yang dicapai (numerik).

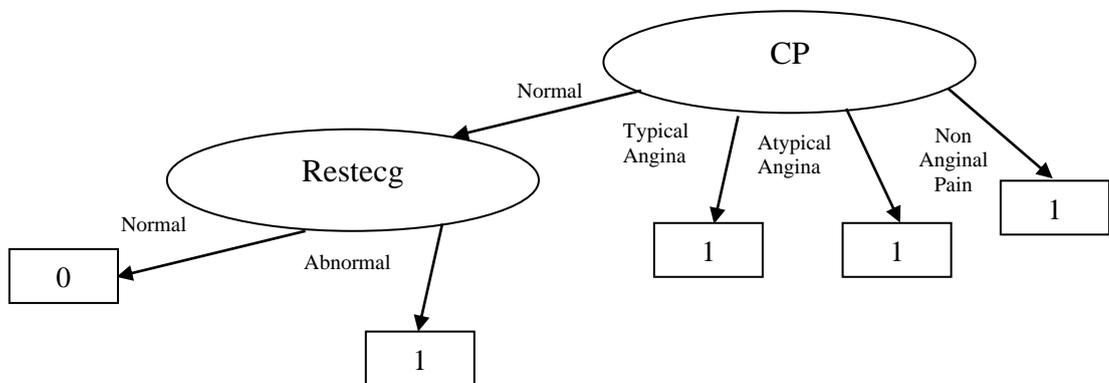
- Exang: Ukuran *boolean* yang menunjukkan apakah latihan angina induksi telah terjadi dengan nilai 1: Ya, 0: Tidak).
- *Slope*: kemiringan segmen ST untuk latihan maksimum (puncak). Terdapat tiga jenis nilai yaitu 1: *upsloping*, 2: *flat*, 3: *downsloping*.
- Target/Klasifikasi dengan nilai *Healthy* (0) atau *Sick* (1).

Selanjutnya dilakukan proses perhitungan *Entropy* dan *Information Gain*.

| Atribut  | Nilai Atribut                        | Total | Entropy  |             | Gain        |                              | Information Gain  | Tertinggi         |
|----------|--------------------------------------|-------|--|-------------|-------------|------------------------------|---|-------------------|
| Total    | Total                                | 10    | $= -3/10 * \log_2(3/10) - 7/10 * \log_2(7/10)$ | 0.881290899 |             |                              |   |                   |
| Age      | 20-50                                | 2     | $= -2/2 * \log_2(2/2)$                         | 0           | 0           | $(2/10 * 0)$                 | $= 0.88129089929852 - (2/10 * 0) - (8/10 * 0.954434002998422)$  | 0.117743696899783 |
|          | >50                                  | 8     | $= -3/8 * \log_2(3/8) - 5/8 * \log_2(5/8)$     | 0.954434003 | 0.763547202 | $(8/10 * 0.954434002998422)$ |   |                   |
| Sex      | Laki-Laki (1)                        | 6     | $= -2/6 * \log_2(2/6) - 4/6 * \log_2(4/6)$     | 0.918295834 | 0.5509775   | $(6/10 * 0.918295834125165)$ | $= 0.88129089929852 - (6/10 * 0.918295834125165) - (4/10 * 0.811278124521573)$  | 0.005802149014792 |
|          | Perempuan (0)                        | 4     | $= -1/4 * \log_2(1/4) - 3/4 * \log_2(3/4)$     | 0.811278125 | 0.32451125  | $(4/10 * 0.811278124521573)$ |   |                   |
| CP       | Normal (0)                           | 5     | $= -3/5 * \log_2(3/5) - 2/5 * \log_2(2/5)$     | 0.970950595 | 0.485475297 | $(5/10 * 0.970950594529396)$ | $= 0.88129089929852 - (5/10 * 0.970950594529396) - (3/10 * 0) - (1/10 * 0) - (1/10 * 0) - (0/10 * 0)$                 | 0.395815602033822 |
|          | Typical Angina (1)                   | 3     | $= -3/3 * \log_2(3/3)$                         | 0           | 0           | $(3/10 * 0)$                 |   |                   |
|          | Atypical Angina (2)                  | 1     | $= -1/1 * \log_2(1/1)$                         | 0           | 0           | $(1/10 * 0)$                 |   |                   |
|          | Non Anginal Pain (3)                 | 1     | $= -1/1 * \log_2(1/1)$                         | 0           | 0           | $(1/10 * 0)$                 |   |                   |
|          | Asymptomatic (4)                     | 0     | 0  | 0           | 0           | $(0/10 * 0)$                 |   |                   |
| Trestbps | Normal (<120) (0)                    | 0     | 0  | 0           | 0           | $(0/10 * 0)$                 | $= 0.88129089929852 - (0/10 * 0) - (5/10 * 0.721928094942925) - (5/10 * 0.721928094942925) - (0/10 * 0) - (0/10 * 0)$ | 0.03485155456236  |
|          | Pra Hipertensi (120-139) (1)         | 5     | $= -1/5 * \log_2(1/5) - 4/5 * \log_2(4/5)$     | 0.721928095 | 0.360964047 | $(5/10 * 0.721928094942925)$ |   |                   |
|          | Hipertensi Peringkat I (140-160) (2) | 5     | $= -2/5 * \log_2(2/5) - 3/5 * \log_2(3/5)$     | 0.970950595 | 0.485475297 | $(5/10 * 0.970950594529396)$ |   |                   |
|          | Hipertensi Peringkat II >            | 0     | 0  | 0           | 0           | $(0/10 * 0)$                 |   |                   |

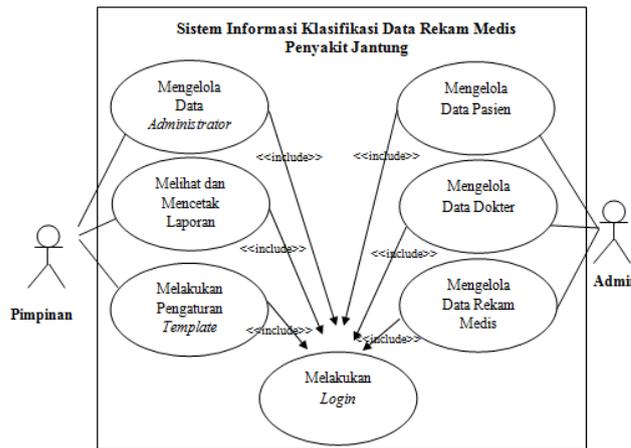
Gambar 2. Perhitungan *Entropy* dan *Information Gain*

Pada gambar 2, nilai dengan *Information Gain* tertinggi diperoleh oleh atribut CP sehingga atribut tersebut akan dijadikan sebagai akar di *node* pertama. Data rekam medis dan atribut yang sudah dijadikan *root* tidak akan diikuti lagi pada iterasi berikutnya. Proses perhitungan *Entropy* dan *Information* dilanjutkan terus hingga seluruh atribut memiliki nilai *Entropy* = 0 yang artinya tidak ada percabangan. Berikut ini hasil pohon keputusan yang terbentuk dari contoh data *training* yang digunakan dapat dilihat pada gambar 1.



Gambar 3. Hasil Pohon Keputusan Yang Terbentuk

- Analisis sistem usulan yaitu menggambarkan sistem usulan yang akan dibangun beserta fitur-fitur yang tersedia yang dimodelkan dengan menggunakan *tools* pemodelan sistem *Use Case Diagram* seperti ditunjukkan oleh gambar 2 berikut.

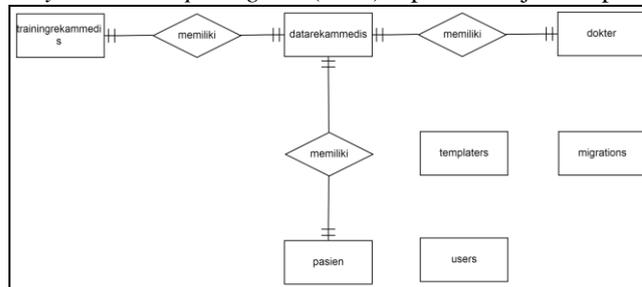


Gambar 4. Use Case Diagram Sistem Usulan

**Perancangan Sistem**

Perancangan sistem pada penelitian ini terbagi menjadi 2 tahapan yaitu:

- Perancangan *prototype* tampilan dengan menggunakan *software* Balsamiq Mockup 3.
- Perancangan basis data yang menunjukkan relasi antar tabel dari setiap basis data yang dimodelkan dengan menggunakan *tools* Entity Relationship Diagram (ERD) seperti ditunjukkan pada gambar 3 berikut.



Gambar 5. Entity Relationship Diagram (ERD) Sistem Usulan

**Pembangunan Sistem**

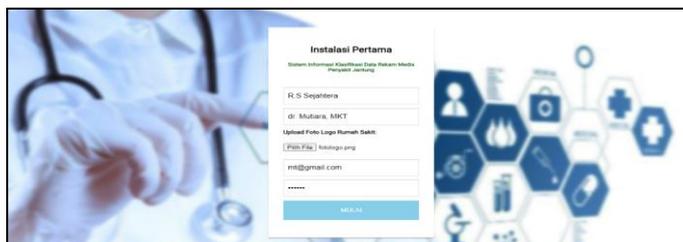
Pembangunan sistem pada penelitian ini menggunakan 2 buah *Framework* yaitu *Framework* Laravel sebagai *Back End* dan *Framework* Bootstrap 3 sebagai *Front End*.

**Pengujian Sistem**

Pengujian sistem pada penelitian ini menggunakan *Confusion Matrix* untuk menguji keakuratan dari algoritma *Decision Tree* C4.5 dalam melakukan klasifikasi.

**4. HASIL DAN PEMBAHASAN**

Pada tahap ini akan ditunjukkan hasil penelitian berupa dibangunnya sebuah sistem informasi klasifikasi data rekam medis penyakit jantung. Berikut ini akan disajikan keseluruhan hasil tampilan dari sistem usulan yang akan dibangun. Tampilan paling awal adalah berupa tampilan untuk melakukan proses instalasi awal. Gambar 6 menunjukkan hasil tampilan instalasi awal.



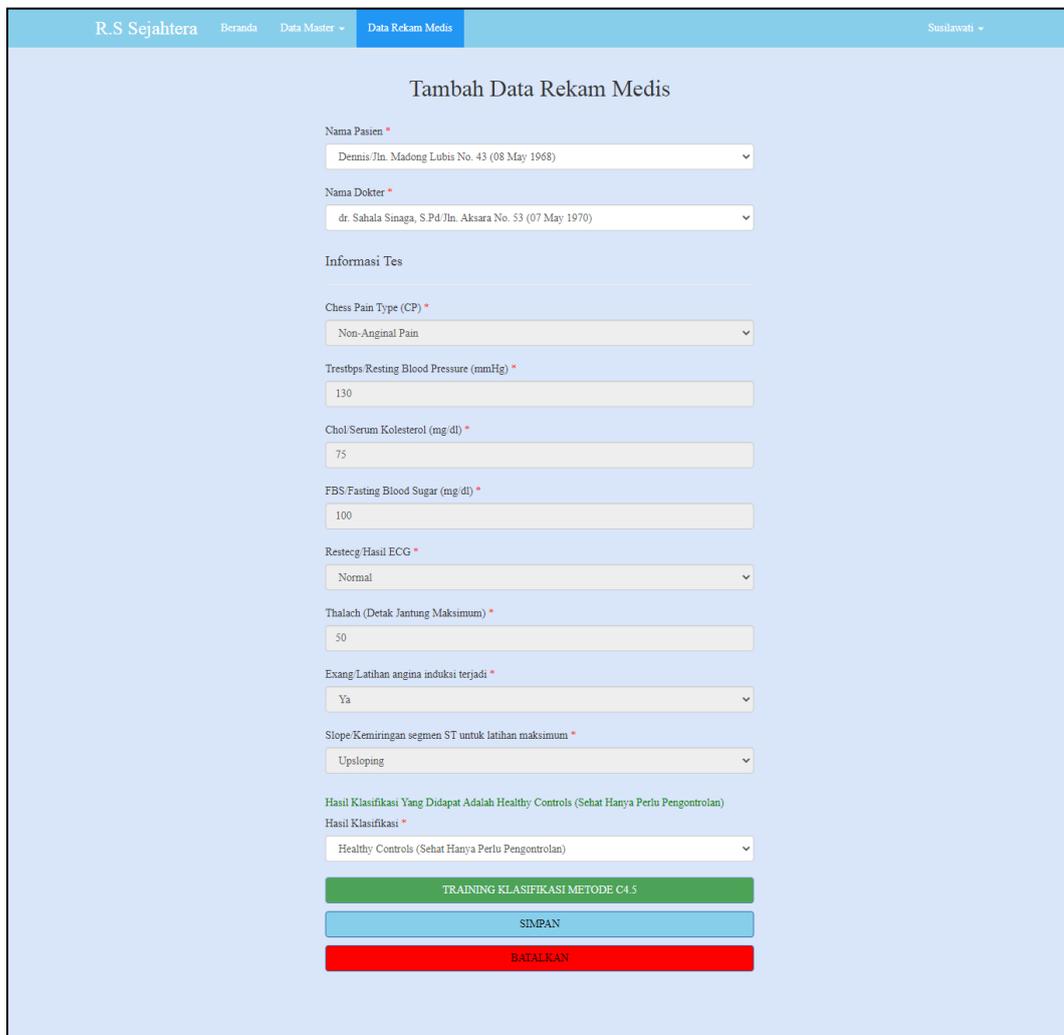
Gambar 6. Tampilan Instalasi Awal

Jika semua *form* telah terisi, maka pengguna cukup menekan tombol mulai untuk memproses *form* yang diisi dan mendaftarkan informasi rumah sakit ke dalam basis data sistem. Setelah proses instalasi selesai dilakukan, maka selanjutnya akan ditampilkan tampilan *login* awal bagi *user* pimpinan dan admin. Gambar 7 menunjukkan hasil tampilan *login*.



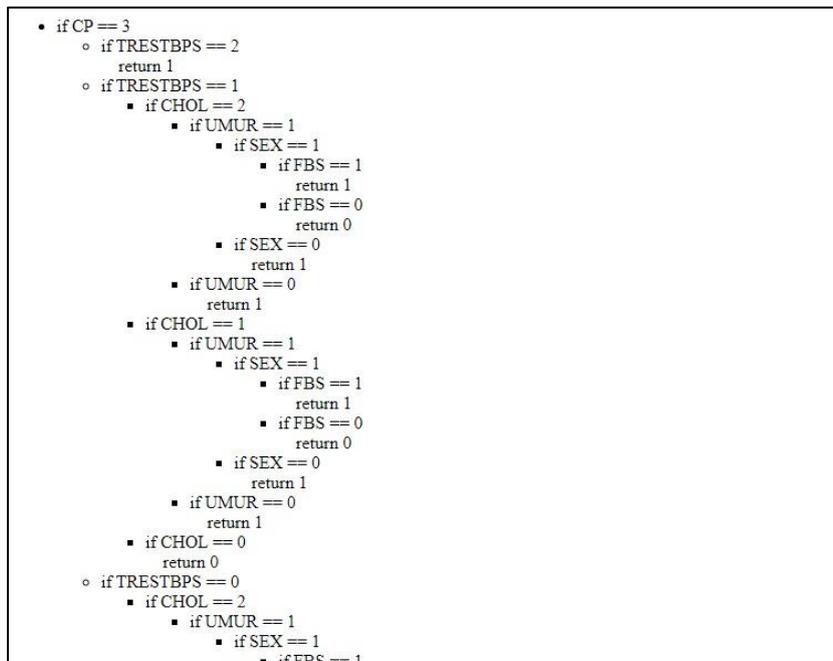
Gambar 7. Tampilan Login

Jika semua *form* telah terisi, maka pengguna cukup menekan tombol proses untuk memproses *form* yang diisikan agar *form* yang diisikan dapat diproses. Setelah *login* berhasil dilakukan, maka pengguna akan ditampilkan tampilan *dashboard* sesuai dengan hak akses mereka. Hak akses pengguna pada sistem yang dibangun terdiri dari dua yaitu administrator dan pimpinan. Implementasi dari algoritma *Decision Tree C4.5* akan dilakukan pada saat penambahan data rekam medis pasien. Gambar 8 akan menunjukkan *form* penambahan data rekam medis pasien penyakit jantung.



Gambar 8. Tampilan Tambah Data Rekam Medis

Ketika tombol *training* klasifikasi metode C4.5 ditekan, maka akan dimunculkan hasil prediksi yang diolah dengan menggunakan algoritma C4.5. Admin boleh menentukan untuk mengikuti hasil prediksi atau tidak. Jika ingin menyimpan data maka tekan tombol simpan. Berikut ini pada gambar 9, akan dilampirkan hasil dari *rule* yang terbentuk berdasarkan data *training* yang diolah dengan algoritma *Decision Tree C4.5*.



Gambar 9. Rule Yang Terbentuk Berdasarkan Data Training Yang Diolah Dengan Algoritma C4.5

Pada tabel 2, akan diuraikan hasil pengujian keakuratan dari algoritma *Decision Tree* C4.5 menggunakan perhitungan secara manual dengan *Confusion Matrix* dimana dataset yang berjumlah 304 data dibagi menjadi 280 data *training* dan 24 data *testing*. Dari 24 data *testing* tersebut terdiri dari 12 data yang diprediksi *sick* dan 12 data yang diprediksi *healthy*. Dari 12 data yang diprediksi *sick* terdapat 2 data yang ternyata diprediksi *healthy* oleh sistem. Kemudian 12 data yang diprediksi *healthy* terdapat 1 data yang ternyata diprediksi *sick* oleh sistem.

Tabel 2. Hasil Pengujian *Confusion Matrix*

| Kelas            | True Sick | True Healthy |
|------------------|-----------|--------------|
| Prediksi Sick    | 12        | 1            |
| Prediksi Healthy | 2         | 11           |

Nilai akurasi dari *Confusion Matrix* tersebut adalah sebagai berikut:

$$\begin{aligned}
 \text{Akurasi} &= \frac{(TN + TP)}{(TN + FN + TP + FP)} \\
 &= \frac{(11 + 12)}{(11 + 1 + 12 + 2)} \\
 &= \frac{23}{24} = 0,96 = 96\%
 \end{aligned}$$

### 5. KESIMPULAN DAN SARAN

Setelah penelitian selesai dilakukan, maka diperoleh kesimpulan sebagai berikut.

- Sistem informasi yang dibangun menerapkan algoritma *Decision Tree* C4.5 dalam melakukan prediksi klasifikasi data rekam medis penyakit jantung.
- Sistem yang dibangun berbasis *website* sehingga memudahkan pengguna dalam mengakses sistem kapanpun dan dimanapun saja asal terhubung dengan internet.
- Berdasarkan keseluruhan hasil pengujian algoritma *Decision Tree* C4.5, algoritma ini tergolong cukup akurat dalam melakukan prediksi dimana hasil pengujian dengan *Confusion Matrix* mendapatkan akurasi sebesar 96%.

### DAFTAR PUSATAKA

[1] E. Kurniawan, "Analisa Data Rekam Medis Menggunakan Teknik Data Mining Association Rules Dengan Algoritma Clustering," in *Seminar Nasional Pendidikan, Sains dan Teknologi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang*, Semarang, 2017.

[2] A. Ali, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering di Rumah Sakit

- Anwar Medika Balong Bendo Sidoarjo," *Jurnal Matrik*, vol. XIX, no. 1, pp. 186-195, 2019.
- [3] Yuliana, "Risiko dan Deteksi Dini Kanker Payudara," *CDK-261*, vol. XLV, no. 2, pp. 144-149, 2018.
- [4] R. T. Handayanto, *Data Mining Dan Machine Learning Menggunakan Matlab Dan Pyth*, Bandung: Informatika, 2020.
- [5] A. Ali, *Penerapan Teknologi Basis Data di Bidang Rekam Medik*, Sidoarjo: Indomedia Pustaka, 2019.
- [6] R. H. A. Alsagheer, A. F. H. Alharan and A. S. A. Al-Haboobi, "Popular Decision Tree Algorithms of Data Mining Techniques: A Review," *International Journal of Computer Science and Mobile Computing*, vol. VI, no. 6, pp. 133-142, 2017.
- [7] E. Elisa, "Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti," *JOIN*, vol. II, no. 1, pp. 36-41, 2017.
- [8] W. I. Rahayu, C. Prianto and E. A. Novia, "Perbandingan Algoritma K-Means dan Naive Bayes Untuk Memprediksi Prioritas Pembayaran Tagihan Rumah Sakit Berdasarkan Tingkat Kepentingan Pada PT. Pertamina (Persero)," *Jurnal Teknik Informatika*, vol. XIII, no. 2, pp. 1-8, 2021.