

KLASIFIKASI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA C4.5

Arni Sepharni¹, Iwansyah Edo Hendrawan², Chaerur Rozikin³
Program Studi Teknik Informatika, Univeritas Singaperbangsa Karawang^{1,2,3}
arnisepharni09@gmail.com

Submitted February 12, 2022; Revised August 8, 2022; Accepted November 29, 2022

Abstrak

Penyakit jantung merupakan salah satu penyakit yang berbahaya. Penyakit jantung dapat membahayakan nyawa penderitanya jika ada keterlambatan dalam penanganannya. Permasalahan ini disebabkan sulitnya deteksi awal pada penderita penyakit jantung karena penderita selalu mengabaikan gejala awal yang timbul. Selain itu biaya yang diperlukan dalam pemeriksaan penyakit jantung tidaklah murah sebab diperlukan pemeriksaan yang dilakukan dokter spesialis serta uji laboratorium. Sistem prediksi merupakan salah satu opsi yang dapat digunakan untuk melakukan deteksi dini pada penderita penyakit jantung dengan biaya yang lebih murah dalam penggunaannya, hal ini disebabkan biaya yang digunakan dalam pemeriksaan dokter spesialis dan tes laboratorium bisa dihilangkan dan digantikan oleh sistem prediksi. Penelitian ini bertujuan untuk membuat sistem prediksi dengan menggunakan algoritma C4.5 dimana algoritma ini membuat prediksi berdasarkan data historis pasien yang akan diperiksa. Hasil yang didapat dari penggunaan algoritma C4.5 untuk melakukan prediksi mendapatkan akurasi 79% sehingga diharapkan bisa menjadi sumber informasi untuk penelitian selanjutnya tentang prediksi dengan menggunakan algoritma C4.5.

Kata Kunci : Penyakit Jantung, Sistem Prediksi, Algoritma C4.5

Abstract

Heart disease is one of dangerous diseases. The late treatment for the disease will endanger the life of the sufferers. This problem is caused by the difficulty of early detection in patients with heart disease because they always ignore the initial symptoms. In addition, the costs involved in examining heart disease are not cheap because it requires examinations carried out by specialists and laboratory tests. The prediction system is one of options that can be used to perform early detection of heart disease patients at a lower cost, with the costs used in examinations by specialist doctors and laboratory tests can be eliminated and replaced by a prediction system. This study aims to create a prediction system using the C4.5 algorithm by which the predictions based on historical data of the patient to be examined are made. The results obtained from using the C4.5 algorithm to make predictions show an accuracy of 79%, meaning that the results can be expected to be a source of information for further research on prediction systems using the C4.5 algorithm.

Key Words : Heart Disease, Prediction System, C4.5 Algorithm

1. PENDAHULUAN

Penyakit jantung adalah sebuah kondisi dimana adanya disfungsi pada kerja jantung. Penyakit pada jantung banyak jenisnya seperti kardiovaskuler, jantung koroner dan serangan jantung [1]. Penyakit jantung disebut oleh WHO sebagai salah satu penyebab penyebab kematian 7,4 juta jiwa pada tahun 2017 [2]. Penyakit jantung dipicu oleh beberapa faktor diantaranya adalah pola hidup yang tidak sehat,

kebiasaan merokok, begadang dan pola makan yang tidak baik [3]. Untuk menyembuhkan penyakit jantung terdapat banyak cara diantaranya adalah seperti operasi, penyinaran dan khemoterapi [4]. Namun masalah yang timbul dari penyakit jantung adalah sulitnya deteksi awal sehingga penderita sadar akan penyakit jantung ketika sudah parah. Deteksi penyakit jantung dapat dilakukan dengan cara konvensional seperti konsultasi dengan dokter spesialis jantung dan tes

laboratorium sehingga biaya yang harus dikeluarkan relatif besar [5].

Kemajuan teknologi telah membuat perkembangan dalam ilmu pengetahuan dan teknologi informasi telah sangat pesat bahkan merambah banyak bidang dan salah satu manfaatnya adalah terciptanya sistem prediksi [6]-[7]. Dunia medis salah satu yang mendapat manfaat perkembangan teknologi dalam penerapan sistem prediksi [8]. Berbagai metode klasifikasi banyak diterapkan untuk melakukan prediksi penyakit [9]. Sistem prediksi sangat berguna dalam memperkirakan sesuatu yang kemungkinan akan terjadi berdasarkan data yang ada sehingga kerugian yang timbul dapat ditanggulangi atau keuntungan yang kemungkinan terjadi dapat dimaksimalkan hasilnya.

Algoritma C4.5 adalah algoritma yang tergolong sebagai *supervised learning* [10]. Algoritma C4.5 sangat baik dalam melakukan klasifikasi dimana klasifikasi menggunakan pohon keputusan (*decision tree*) [11]. Bentuk pohon keputusan mirip *flowchart* yang disetiap simpul internal menguji tiap atribut dan diantara cabangnya adalah sebuah hasil dan di tiap simpul daun atau terminal adalah label kelas [12]. Algoritma C4.5 menggunakan konsep *gain* atau *entropy reduction* untuk membuat pembagian yang optimal serta dapat membuat aturan-aturan yang bisa diintegrasikan dan paling cepat dibanding algoritma lainnya. yang hasilnya adalah sebuah pohon keputusan [6] [13] [14].

Tujuan penelitian ini adalah untuk melihat efektifitas algoritma C4.5 dalam melakukan prediksi penyakit jantung. Dalam hal ini akurasi yang akan dibuat menjadi acuan dalam penilaian performa algoritma dalam prediksi penyakit jantung.

2. METODE PENELITIAN

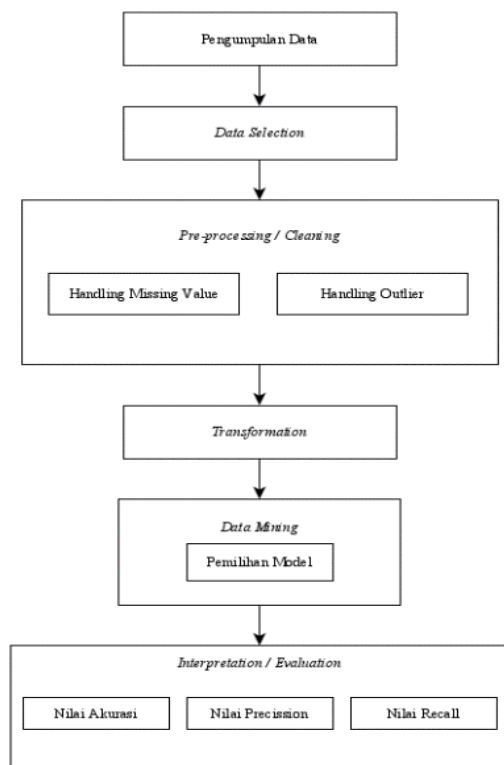
Metodologi dari penelitian ini dilakukan dengan tahapan teknik pengumpulan data dan beberapa tahapan yang ada didalam teknik pemrosesan data.

Teknik Pengumpulan Data

Data yang diperoleh menggunakan teknik pengambilan data sekunder. Data diambil dari situs *Kaggle Dataset* yang dapat diakses melalui alamat *website* <https://www.kaggle.com/fedesoriano/heart-failure-prediction>. Dataset tersebut berisi gejala-gejala pada penyakit jantung.

Teknik Pemrosesan Data

Pemrosesan data menggunakan *tools Google Colab* dan juga menggunakan bahasa pemrograman *python*. Pengolahan data pada penelitian ini menggunakan metode *Knowledge Discovery in Database (KDD)*. *Knowledge Discovery in Database (KDD)* merupakan metode yang dilakukan untuk mendapatkan informasi atau pengetahuan dari basis data yang tersembunyi [15]. Hasil penelitian tersebut nantinya akan digunakan dalam pengambilan keputusan sebagai basis pengetahuan (*knowledge base*) [16]. Berikut adalah tahapan atau proses dalam penelitian ini pada Gambar 1.



Gambar 1. Tahapan KDD

a. Data Selection

Data selection atau seleksi data merupakan tahapan yang dilakukan untuk memilih data dari basis data operasional [17]. Penentuan basis data yang akan digunakan untuk prediksi merupakan tahapan pertama dalam *data mining*. Data yang akan digunakan adalah data yang berasal dari data *Kaggle Dataset*. Pada tahap seleksi data ini, data yang berasal dari basis data akan diseleksi, sehingga data yang sudah diseleksi tersebut akan dianalisis dan masuk ke tahap selanjutnya.

b. Preprocessing / Cleaning

Preprocessing merupakan salah satu tahapan dalam proses *data mining* yang akan digunakan dalam melakukan *cleaning* atau pembersihan pada *dataset* yang memiliki *noise* [18]. Tahap *preprocessing* ini dilakukan untuk menangani data-data yang tidak diperlukan, seperti data yang memiliki *missing value*, *outlier*, maupun *imbalance data*. Data yang memiliki *missing value*

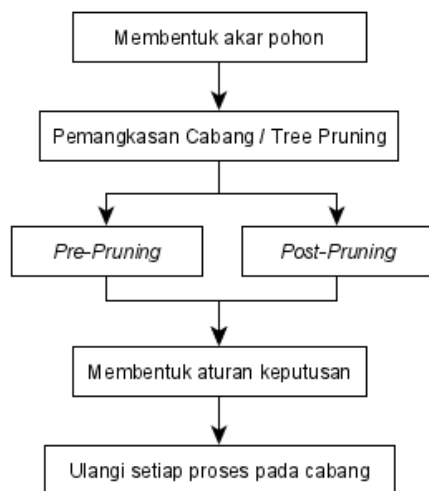
ataupun *outliers* tersebut nantinya akan dilakukan penanganan dengan cara menghapus sejumlah *record*. Sedangkan penanganan untuk *imbalance data* akan dilakukan proses pendekatan *re-sampling*.

c. Transformation

Pada tahapan transformasi data ini dilakukan perubahan terhadap data kedalam bentuk data yang sesuai pada proses *data mining*. Adapun teknik yang dapat dilakukan pada proses *transformation data* yaitu teknik *descretize*. Teknik *descretize* merupakan teknik yang digunakan untuk mengubah tipe data kontinu menjadi tipe data diskrit. Selain itu *descretize* juga dapat digunakan dalam mengurangi jumlah atribut numerik dengan membuat label interval dan membagi range atribut tersebut [19].

d. Data Mining

Data yang sudah melakukan tahap *preprocessing* dan *transformation data*, maka langkah selanjutnya akan dilakukan pengolahan data dengan menggunakan metode atau algoritma *data mining*. Pada penelitian ini menggunakan algoritma C4.5 untuk proses analisis gejala penyakit jantung. Hasil dari analisis gejala penyakit jantung dengan menggunakan algoritma C4.5 tersebut akan menghasilkan pohon keputusan yang dapat memprediksi penyakit jantung berdasarkan *rule* dari pohon keputusan tersebut. Menurut [20] adapun tahapan dalam membuat sebuah pohon keputusan dengan algoritma C4.5 dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Algoritme C4.5

1. Tahapan pertama sebelum membuat pohon keputusan yaitu adalah dengan membentuk akar pohon terlebih dahulu, setelah itu data akan dibedakan dengan atribut yang sesuai untuk membentuk daun.
2. Tahap selanjutnya dilakukan pemangkasan pada cabang yang tidak diperlukan atau biasa disebut dengan *tree pruning*. Pemangkasan cabang ini dilakukan penyederhanaan pada pohon keputusan serta memangkas sejumlah ekor yang dihasilkan dari hasil prediksi. Ada dua tahap dalam proses dua tahap dalam proses *tree pruning* yaitu *pre-pruning* dan *post-pruning*.
 - a. *Pre-pruning* adalah proses yang dilakukan dalam pembentukan pada *subtree* agar dihentikan lebih awal. Pada saat berhenti *node* akan berubah menjadi *node* akhir.
 - b. *Post-pruning* merupakan tahapan untuk melakukan penyederhanaan atau pemangkasan pada cabang-cabang *subtree* yang telah dibentuk. *Node* yang jarang dipotong akan menjadi *node* akhir.

3. Membuat aturan keputusan dari akar sampai daun dari pohon yang sudah dibentuk sebelumnya.
4. Ulangi setiap proses yang dilakukan pada setiap cabang sampai menemukan kelas yang sama yang nantinya akan dijadikan sebagai akar yang disesuaikan dari nilai gain tertinggi dari atribut-atribut yang ada.

e. Interpretation / Evaluation

Evaluasi yang dihasilkan dari model yang dibuat dengan menggunakan algoritma C4.5 berupa pohon keputusan dapat menghasilkan nilai akurasi, *precision*, dan *recall* dari pengujian menggunakan *decision tree classifier* dan pembagian data menggunakan *k-fold cross validation*. Adapaun *interpretation* dari tahap ini yaitu menampilkan pola informasi yang dapat dipahami bagi pengguna yang memakainya.

3. HASIL DAN PEMBAHASAN

Pada proses sebelumnya telah dijelaskan beberapa tahapan sebelum melakukan proses *data mining* seperti *data selection*, *preprocessing*, *transformation*, dan *evaluation*. Pada penelitian ini penerapan *data mining* dengan menggunakan algoritma C4.5 untuk memprediksi gejala penyakit jantung dengan hasil akhir berupa *rules* yang didapatkan dari pohon keputusan berupa nilai akurasi, *precision*, dan *recall*.

Pengumpulan Data

Data diperoleh dari situs kaggle yang terdiri dari beberapa atribut dan label diantaranya *Age*, *Sex*, *Chestpaine*, *RestingBP*, *Cholesterol*, *FastingBS*, *RestingECG*, *MaxHR*, *ExerciseAngina*, *Oldpeak*, *STSLope*, dan *Heart Disease* sebagai label. Deskripsi data tersebut dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Data

No	Atribut	Tipe Data	Keterangan
1	Age	Numerik	Usia Pasien
2	Sex	Kategorik	Jantung Jenis Kelamin pada pasien jantung
3	Chestpaintype	Kategori	Tipe sakit dada pada pasien jantung
4	Resting BP	Numerik	Tekanan Darah pada pasien jantung
5	Cholesterol	Numerik	Kolesterol pada pasien jantung
6	Fasting BS	Numerik	-
7	Resting ECG	Kategorik	-
8	Max HR	Numerik	Detak Jantung pada pasien jantung
9	Exercise Angina	Kategorik	Rasa sakit didada pada pasien jantung
10	Oldpeak	Numerik	Tipe depresi pada pasien jantung
11	ST Slope	Kategorik	Kemiringan detak jantung pada pasien jantung diukur menggunakan alat EKG
12	Heart Disease	Numerik	Label keterangan pasien tersebut menderita penyakit jantung atau tidak

Seleksi Data

Seleksi data dilakukan karena tidak semua atribut yang terdapat pada data digunakan untuk proses pengolahan *data mining*. Proses pada seleksi data tersebut dilakukan untuk memilih atribut atau variabel yang berpengaruh pada data gejala penyakit jantung untuk memprediksi penyakit jantung. Atribut pada data yang dipilih dan akan diolah terlihat pada Tabel 2.

Tabel 2. Atribut Terpilih

No	Atribut Terpilih	Detail Penggunaan
1	Resting BP	Atribut Penjelas
2	Cholesterol	Atribut Penjelas
3	Max HR	Atribut Penjelas
4	Exercise Angina	Atribut Penjelas
5	ST Slope	Atribut Penjelas
6	Heart Disease	Atribut Target / Kelas Target

Atribut *RestingBP*, *Cholesterol*, *MaxHR*, *ExerciseAngina*, *STSlope*, dan *Heart Disease* dipilih karena berpengaruh terhadap gejala-gejala yang paling dialami oleh penderita penyakit jantung. Sedangkan untuk atribut yang tidak digunakan akan dihapus. Data yang sudah terpilih tersebut akan diolah dengan menggunakan algoritma C4.5 yang nantinya akan menghasilkan sebuah *rules* yang didapatkan dari pohon keputusan yang dapat digunakan untuk memprediksi penyakit jantung.

Preprocessing Data

Preprocessing atau pembersihan data merupakan tahapan penting dalam proses *mining* dikarenakan tidak selamanya data yang digunakan pada proses *mining* tersebut dalam kondisi yang bersih. Selain itu, data yang masih memiliki permasalahan nantinya akan berpengaruh terhadap hasil yang didapatkan dari proses *mining*. Ada beberapa data yang masih harus dilakukan penanganan khusus seperti menangani *missing value* dan *outlier*. Setelah melakukan analisis pada tahapan *data preprocessing* sebelumnya terdapat beberapa data yang harus dilakukan penanganan terhadap *missing value* dan *outliers* pada data.

1. Penanganan *Outliers Data*

Pada *dataset* yang sudah diseleksi, ada beberapa data yang memiliki *outliers* sehingga perlu ditangani. *Outliers* dalam *dataset* dapat dilihat pada Tabel 3.

Tabel 3. *Outliers Data*

No	Nama Atribut	Min	Max	Average
1	Resting BP	0	200	132.4
2	Cholesterol	0	603	198.8

Penanganan pada *outliers* data tersebut adalah dengan menghapus *range* yang terlalu jauh, seperti *max value* pada atribut *restingBP* dan *cholesterol* yang bernilai 200 - 603, lalu atribut yang melebihi *range* tersebut akan dihapus.

2. Penanganan *Missing Value*

Pada *dataset* yang sudah diseleksi, ternyata tidak memiliki *missing value* pada atribut *RestingBP*, *Cholesterol*, *MaxHR*, *ExerciseAngina*, *STSlope*, dan *Heart Disease* dapat dilihat pada keterangan Gambar 2.

```
[ ] ds.isnull().any()
#Mengecek adakah nilai null/missing

Age           False
Sex           False
RestingBP     False
Cholesterol   False
MaxHR         False
ExerciseAngina False
ST_Slope      False
HeartDisease  False
dtype: bool
```

Gambar 3. Missing Value

Dilihat pada gambar diatas bahwa data tersebut tidak memiliki *missing value*.

a. Transformation Data

Transformation data merupakan tahapan yang dilakukan untuk mengubah data ke dalam bentuk yang sesuai dengan kebutuhan yang ada. Pada *dataset* data yang akan dilakukan transformasi yaitu pada atribut *ExerciseAngina* dan *STSlope* dikarenakan pada proses evaluasi yang dilakukan pada *python* atribut tersebut harus menggunakan tipe data numerik.

Tabel 4. Perubahan Data *Exercise Angina*

No	Atribut <i>Exercise Angina</i>	Detail Penggunaan
1	Tidak	0
2	Ya	1

Perubahan tersebut akan diterapkan pada *dataset* yang memiliki *class* “Tidak” diganti menjadi nilai 0, dan *class* “Ya” diganti menjadi nilai 1.

Tabel 5. Perubahan Data Atribut *ST Slope*

No	Atribut <i>ST Slope</i>	Detail Penggunaan
1	Up	0
2	Flat	1
3	Down	2

Perubahan tersebut akan diterapkan pada *dataset* yang memiliki *class* “Up” diganti menjadi nilai 0, dan *class* “Flat” diganti

menjadi nilai 1, dan *class* “Down” diganti menjadi nilai 2.

b. Data Mining

Pada tahap sebelumnya telah dilakukan beberapa tahapan diantaranya *data selection*, *preprocessing*, dan *transformation*. Teknik pemodelan data *mining* yang dipilih adalah klasifikasi prediksi dengan menggunakan algoritma C4.5. Klasifikasi algoritma C4.5 digunakan untuk mencapai tujuan awal penelitian yaitu memprediksi penyakit jantung.

Berikut adalah rumus perhitungan yang terdapat pada *Decision Tree* yaitu *Entropy* dan *Gain*.

$$Entropy(S) = -\sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Implementasi algoritma C4.5 menggunakan bahasa pemrograman *python* dengan bantuan IDE atau *Google Colab* guna membuat model klasifikasi dengan metode C4.5 dan juga dapat memprediksi penyakit jantung. Model yang dibuat menggunakan *Decision tree classifier*, sebuah library dari *sklearn* yang dapat memproses data menggunakan algoritma *decision tree*, lalu data akan dibagi menjadi data uji dan data latih. Pembagian data uji dan data latih tersebut menggunakan *k-folds cross validation* dengan nilai k=10, sehingga nantinya akan terdapat 10 skenario pengujian dengan data uji dan data latih yang berbeda beda. Proses split data menggunakan *k-folds cross validation* dengan nilai k=10 yang berarti terdapat 10 skenario pengujian. Terdapat 10 skenario pengujian dengan data uji dan data latih yang berbeda-beda. Berikut adalah penjelasan mengenai skenario yang dilakukan. Berikut adalah contoh untuk skenario pertama.

1. Skenario 1

Tabel 6. Hasil Prediksi Skenario 1

	Actual Tidak	Actual Ya
Pred. Tidak	67	16
Pred. Ya	15	86

Berdasarkan hasil pada Tabel 6 terlihat bahwa sebanyak 86 data diprediksi “Ya” dengan valid kemudian 67 data diklasifikasikan “Tidak” dengan valid. Setelah diketahui hasil dari prediksi tersebut, maka dapat diketahui nilai *accuracy*, *recall*, *precision* dan *f1score*.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$= \frac{86 + 67}{184} = 0.8315$$

Precision, recall dan f1score class “Ya” pada skenario 1

$$precision = \frac{TP}{TP + FP} = \frac{86}{86 + 16} = 0,8431$$

$$recall = \frac{TP}{TP + FN} = \frac{86}{86 + 15} = 0,8514$$

Precision, recall dan f1score class “Tidak” pada skenario 1

$$precision = \frac{TN}{TN + FP} = \frac{67}{67 + 16} = 0,8072$$

$$recall = \frac{TN}{TN + FN} = \frac{67}{67 + 15} = 0,8170$$

Dalam skenario pertama ini mendapatkan akurasi yang baik, jika dapat memprediksi semuanya maka *precision* dan *recall* pun akan bernilai sama. Terdapat 10 skenario pengujian yang dilakukan pada penelitian ini, penulis hanya menjelaskan detail pada skenario pertama saja, selanjutnya skenario 2 sampai skenario 10 cara dan rumusnya sama persis dengan skenario pertama namun hasil akurasi, *precision* dan *recall* berbeda karna setiap skenario memiliki data uji dan data latih yang berbeda-beda.

Berikut adalah hasil evaluasi dari 10 skenario pengujian pada **Tabel 7**.

Tabel 7. Hasil Evaluasi

Sk	Akurasi	Precision	Recall
1	0,7608	0,8085	0,7450
2	0,6739	0,7058	0,7058
3	0,8260	0,8723	0,8039
4	0,8260	0,8181	0,8823
5	0,7717	0,812	0,7647
6	0,8043	0,8367	0,8039
7	0,75	0,7692	0,7843
8	0,7065	0,8	0,6274
9	0,7362	0,76	0,76
10	0,7362	0,75	0,78

Keterangan:

SK = Skenario

Terlihat pada Tabel 7, model yang dibuat belum tentu sempurna, setelah melakukan skenario terhadap *cross validation*, maka seluruh proses dari skenario pertama sampai akhir dirata-ratakan tingkat akurasi dari proses *cross validation* tersebut, dan hasil akurasi rata ratanya adalah 0,7592, *precision* mendapatkan skor rata-rata 0,7933, dan *recall* mendapatkan nilai rata-rata 0,7657.

c. Evaluasi

Proses KDD (*Knowledge Discovery in Database*) dilakukan pada penelitian ini dan memiliki beberapa tahapan yaitu, *data selection*, *data processing*, *transformation*, *data mining*, dan *evaluation*. Dari skenario yang sudah dilakukan dengan menggunakan algoritme C4.5 menghasilkan suatu *rules* atau aturan yang didapatkan dari pemodelan pohon keputusan. Pohon keputusan tersebut dapat dilihat pada Gambar 4.



Gambar 4. Pohon Keputusan

Berdasarkan pohon keputusan mengenai penyakit stroke yang sudah dibangun menggunakan algoritma C4.5 tersebut, menghasilkan beberapa aturan atau *rules* diantaranya :

- a. JIKA ST Slope ≤ 1.500 dan Max HR ≤ 183 dan Max HR ≤ 150.500 MAKA tidak
- b. JIKA ST Slope ≤ 1.500 dan Max HR ≤ 183 dan Max HR > 150.500 dan kolesterol ≤ 70.500 MAKA tidak
- c. JIKA ST Slope ≤ 1.500 dan Max HR ≤ 183 dan Max HR > 150.500 dan kolesterol > 70.500 dan resting BP ≤ 156 MAKA stroke
- d. JIKA ST Slope ≤ 1.500 dan Max HR ≤ 183 dan Max HR > 150.500 dan kolesterol > 70.500 dan resting BP > 156 MAKA tidak
- e. JIKA ST Slope ≤ 1.500 dan Max HR > 183 MAKA stroke
- f. JIKA ST Slope > 1.500 dan kolesterol ≤ 42.500 dan max HR ≤ 99.500 MAKA tidak
- g. JIKA ST Slope > 1.500 dan kolesterol ≤ 42.500 dan max HR > 99.500 dan resting BP ≤ 104.500 MAKA stroke
- h. JIKA ST Slope > 1.500 dan kolesterol ≤ 42.500 dan max HR > 99.500 dan resting BP > 104.500 MAKA tidak
- i. JIKA ST Slope > 1.500 dan kolesterol > 42.500 dan exercise angina ≤ 0.500 MAKA stroke
- j. JIKA ST Slope > 1.500 dan kolesterol > 42.500 dan exercise

- angina > 0.500 dan kolesterol ≤ 197 MAKA tidak
- k. JIKA ST Slope > 1.500 dan kolesterol > 42.500 dan exercise angina > 0.500 dan kolesterol > 197 MAKA stroke

4. SIMPULAN

Berdasarkan penelitian yang telah dilakukan, maka dapat disimpulkan bahwa algoritma C4.5 dapat digunakan untuk memprediksi penyakit jantung dengan menggunakan beberapa teknik yang dapat digunakan untuk melihat hasil prediksi seperti *cross validation* dan *confusion matrix*.

Hasil penelitian yang telah dilaksanakan menggunakan algoritma C4.5 diterapkan dengan eksperimen menggunakan *k-fold cross validation* dengan jumlah $k = 10$, maka dari itu pengujian ini memiliki 10 skenario pengujian dan 10 skenario pembagian dataset. Pembagian dataset menghasilkan nilai akurasi, *precision*, *recall* yang berbeda-beda tiap pembagian datanya dengan kata lain pembagian data berpengaruh terhadap hasil dari algoritma C4.5. Performansi algoritma C4.5 diukur menggunakan nilai akurasi, *precision*, *recall*. Pengujian 10 skenario tersebut menghasilkan tingkat akurasi rata-rata yaitu 0,7592, *precision* mendapatkan skor rata-rata 0,7933, *recall* mendapatkan nilai rata-rata 0,7657.

DAFTAR PUSTAKA

- [1] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, pp. 437–444, 2020.
- [2] Y. D. Hastuti and E. D. Mulyani, "Kecemasan Pasien dengan Penyakit Jantung Koroner Paska Percutaneous Coronary

- Intervention,” *J. Perawat Indones.*, vol. 3, no. 3, pp. 167–174, 2019.
- [3] H. M. Nawawi, J. J. Purnama, and A. B. Hikmah, “Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung,” *J. Pilar Nusa Mandiri*, vol. 15, no. 2, pp. 189–194, 2019.
- [4] M. A. Bianto, K. Kusriani, and S. Sudarmawan, “Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, pp. 75–83, 2020.
- [5] A. B. Wibisono and A. Fahrurrozi, “Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner,” *J. Ilm. Teknol. dan Rekayasa*, vol. 24, no. 3, pp. 161–170, 2020.
- [6] W. D. Septiani, “Komparasi Metode Klasifikasi Data Mining Algoritma C4. 5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis,” *J. Pilar Nusa Mandiri*, vol. 13, no. 1, pp. 76–84, 2017.
- [7] H. Husdi and Y. Lasena, “Real Time Analisis Berbasis Internet Of Things Untuk Prediksi Iklim Lahan Pertanian,” *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 3, pp. 834–840, 2020.
- [8] M. B. Priyantono, A. A. Rachmawan, L. A. P. Budi, and K. C. Kirana, “Sistem Prediksi Gejala Virus Korona dengan Metode Forward Chaining,” *JTERA (Jurnal Teknol. Rekayasa)*, vol. 5, no. 1, pp. 111–118, 2020.
- [9] T. Praningki and I. Budi, “Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN,” *Creat. Inf. Technol. J.*, vol. 4, no. 2, pp. 83–93, 2018.
- [10] T. Sutrisno and S. Claudia, “Analisis Rekomendasi Peminatan Menggunakan Metode Decision Tree dengan Algoritma C4. 5,” *Comput. J. Comput. Sci. Inf. Syst.*, vol. 2, no. 2, pp. 178–185, 2018.
- [11] E. Ermawati, “Algoritma Klasifikasi C4. 5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai,” *Sist. J. Sist. Inf.*, vol. 8, no. 3, pp. 513–528, 2019.
- [12] S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019.
- [13] I. P. Astuti, “Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4. 5,” *Fountain Informatics J.*, vol. 2, no. 2, pp. 41–45, 2017.
- [14] R. S. Asa, “Identifikasi Penyaluran Zakat Menggunakan Algoritma C4. 5 (Studi Kasus di BAZNAS Kabupaten Agam),” *J. Sains dan Inform. Res. Sci. Inform.*, vol. 5, no. 1, pp. 44–53, 2019.
- [15] R. T. Prasetio and E. Ripandi, “Optimasi Klasifikasi jenis hutan menggunakan deep learning berbasis optimize selection,” *J. Inform.*, vol. 6, no. 1, pp. 100–106, 2019.
- [16] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, 2017.
- [17] N. Lestari, R. F. Gunawan, S. Informasi, S. Bangek, K. Tangah, and W. Sumatera, “Implementasi Data Mining untuk Menentukan Pola Penjualan dengan Market Basket Analysis,” *Inf. Syst. Res. J.*, vol. 1, 2021.
- [18] K. Fatmawati and A. P. Windarto, “Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit

- Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi,” *CESS (Journal Comput. Eng. Syst. Sci.*, vol. 3, no. 2, pp. 173–178, 2018.
- [19] I. Hidayanti, T. B. Kurniawan, and A. Afriyudi, “Perbandingan Dan Analisis Metode Klasifikasi Untuk Menentukan Konsentrasi Jurusan,” *J. Ilm. Inform. Glob.*, vol. 11, no. 1, pp. 16–21, 2020, doi: 10.36982/jig.v11i1.1067.
- [20] F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.