

# **PENERAPAN DECISION TREE DALAM PENENTUAN INDIKATOR LETAL GAGAL JANTUNG**

**Yuzzar Rizky Mahendra<sup>1)\*</sup>, Sitti Ayuningrum<sup>1)</sup>, Rizki Dwi Rahmawan<sup>1)</sup>**

**<sup>1</sup>Informatika, Fakultas Teknologi Informasi dan Sains Data, Universitas  
Sebelas Maret, Surakarta**

**\*Penulis korespondensi: yuzzarrizkymahendra@student.uns.ac.id**

## **ABSTRAK**

Gagal jantung merupakan masalah kesehatan yang progresif dengan angka mortalitas dan morbiditas yang tinggi di negara maju maupun negara berkembang, termasuk Indonesia. Oleh karena itu, bantuan dalam memprediksi penyakit gagal jantung sangat dibutuhkan dalam mendeteksi dan menentukan indikator yang dapat mengarah kepada penyakit gagal jantung yang letal. Jurnal ini bertujuan untuk memberikan informasi terkait indikator letal pada penyakit gagal jantung. Metode yang digunakan merupakan pohon keputusan dimana indikator-indikator yang dimiliki pasien diolah untuk mendapatkan indikator letal terhadap penyakit gagal jantung. Diharapkan jurnal ini dapat membantu tenaga medis dalam mendiagnosis pasien sehingga kemungkinan penyakit gagal jantung letal dapat dicegah.

**Kata-kata kunci:** Gagal Jantung, Indikator Letal, Pohon Keputusan, Letal

## **ABSTRACT**

*Heart failure is a progressive health problem with high mortality and morbidity rates in both developed and developing countries, including Indonesia. Therefore, assistance in predicting heart failure is urgently needed in detecting and determining indicators that can lead to lethal heart failure. This journal aims to provide information related to lethal indicators in heart failure. The method used is a decision tree where the indicators owned by patients are processed to obtain lethal indicators for heart failure. It is hoped that this journal can assist medical personnel in diagnosing patients so that the possibility of lethal heart failure can be prevented.*

**Keywords:** Heart Failure, Lethal Indicator, Decision Tree, Lethal.

## Pendahuluan

Istilah umum untuk semua jenis gangguan yang mempengaruhi jantung disebut penyakit jantung. Penyakit jantung adalah penyakit yang hanya mengacu pada jantung, sedangkan penyakit kardiovaskular mengacu pada gangguan pembuluh darah dan jantung. Penyakit kardiovaskular (CVD) adalah penyebab utama kematian global, mempengaruhi estimasi 17,9 juta nyawa setiap tahunnya. Faktor risiko perilaku yang paling penting dari penyakit jantung dan stroke adalah pola makan yang tidak sehat, ketidakaktifan fisik, penggunaan tembakau, dan penggunaan alkohol yang tidak sehat. Efek dari faktor risiko yang mungkin muncul pada setiap individu antara lain, meningkatnya tekanan darah, meningkatnya gula darah, meningkatnya lemak darah, kelebihan berat badan, dan obesitas. Faktor-faktor risiko ini dapat diukur di fasilitas kesehatan primer dan mengindikasikan peningkatan risiko untuk penyakit jantung, gagal jantung, dan komplikasi lainnya. (World Health Organization 2021)

Gagal jantung merupakan masalah kesehatan yang progresif dengan angka mortalitas dan morbiditas yang tinggi di negara maju maupun negara berkembang termasuk di Indonesia. Di Indonesia sendiri, pasien gagal jantung relatif muda dibandingkan dengan negara di Eropa dan Amerika. Prevalensi dari gagal jantung sendiri semakin meningkat karena pasien yang mengalami kerusakan jantung yang bersifat akut dapat berlanjut menjadi gagal jantung kronik (Siswanto et al. 2015). Penentuan indikator letal pada gagal jantung sangat penting untuk mengidentifikasi pasien berisiko tinggi dan menerapkan langkah-langkah penanganan yang tepat.

Dalam beberapa tahun terakhir, perkembangan teknologi dan kecerdasan buatan telah membuka peluang baru dalam penggunaan metode yang efektif dalam bidang kesehatan. Salah satu metode yang digunakan adalah *Decision Tree* atau pohon keputusan. *Decision tree* adalah algoritma pembelajaran mesin yang populer dan efisien dalam pengklasifikasian dan prediksi. Klasifikasi adalah jenis analisis data yang dapat membantu orang memprediksi label kelas sampel yang harus diklasifikasikan. Berbagai macam teknik klasifikasi telah diusulkan dalam bidang-bidang seperti mesin, sistem pakar dan statistik.

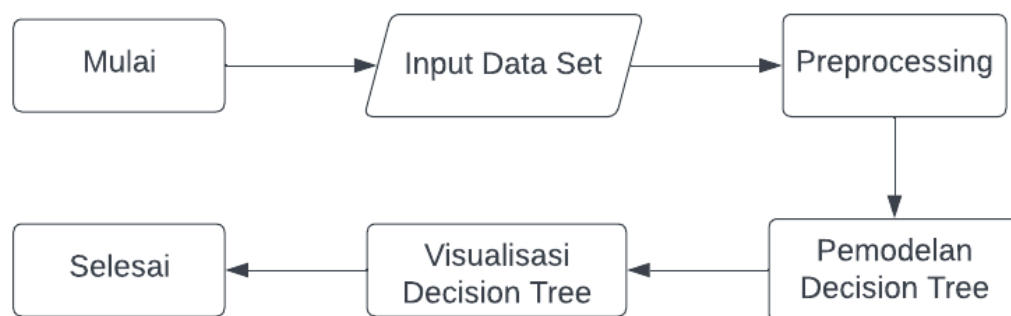
Pada pengklasifikasian menggunakan decision tree, pohon (tree) adalah struktur data yang terdiri dari simpul (node) dan rusuk (edge). Simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root node), simpul percabangan/internal (branch/internal node), dan simpul daun (leaf node) (Hermawati F. A. 2013). Dalam konteks penentuan indikator letal gagal jantung, penerapan *Decision Tree* dapat memberikan pandangan yang lebih dalam mengenai faktor-faktor risiko dan variabel-variabel yang berperan penting dalam memprediksi tingkat kematian pasien.

Penelitian sebelumnya telah mengaplikasikan *Decision Tree* dalam penentuan indikator letal gagal jantung. Misalnya, menurut Suen (2019) menggunakan pendekatan *Decision Tree* untuk memprediksi mortalitas pasien gagal jantung selama masa rawat inap. Hasil

penelitian mereka menunjukkan bahwa *Decision Tree* dapat menjadi alat yang efektif dalam menentukan indikator letal gagal jantung.

Namun, masih terdapat ruang untuk penelitian lebih lanjut dalam menerapkan *decision tree* dalam konteks penentuan indikator letal gagal jantung. Penelitian ini bertujuan untuk menerapkan *decision tree* dalam menentukan indikator letal gagal jantung dan juga membantu tenaga kesehatan dalam memprediksi kemungkinan gagal jantung yang letal. Diharapkan hasil dari penelitian ini dapat memberikan wawasan baru dalam penentuan indikator letal gagal jantung dan meningkatkan pengelolaan pasien dengan kondisi tersebut.

## Metode



Gambar 1. Flowchart Metode Pengolahan Data

Dalam jurnal ini, kita menggunakan metode *Decision Tree* dengan algoritma *CART*. Aplikasi yang digunakan untuk mengolah data set menggunakan *Google Collab*. Diagram di atas adalah gambar 1 skema proses pengolahan data yang akan kita lakukan. Proses pengolahan akan dijelaskan lebih lanjut seperti berikut:

### 1. Input Data Set

Data set yang akan kita gunakan, kita ambil dari *UCI Machine Learning* dengan judul “*Heart failure clinical records Data Set*”. Kita akan menginputkan data tersebut ke dalam *Google Collab* agar nantinya kita dapat melakukan proses klasifikasi *Decision Tree*.

### 2. Preprocessing

Pada penelitian ini, proses *preprocessing* akan melakukan hal-hal seperti berikut:

#### 1) Memisahkan Nilai atribut dan Nilai tujuan

Jurnal ini menggunakan dataset dengan 13 kolom yaitu *Age*, *Anemia*, *High Blood Pressure*, *Creatinine Phosphokinase*, *Diabetes*, *Ejection Fraction*, *Platelets*, *Sex*, *Serum Creatine*, *Serum Sodium*, *Smoking*, *Time*, dan *Death*

*Event*. Dengan Kolom Death Event dikelompokkan menjadi 2 yaitu boolean bernilai 1 yang berarti meninggal, dan boolean bernilai 0 yang berarti tidak meninggal dalam masa *follow-up*. Selain Death Event, Anemia, Diabetes, High Blood Pressure, Sex, dan Smoking menggunakan boolean (angka 1 dan 0).

Tabel 1  
Informasi Atribut

No	Atribut	Penjelasan
1	<i>Age</i>	Usia pasien (Tahun)
2	<i>Anemia</i>	Penurunan sel darah merah atau hemoglobin (boolean: 1 artinya iya dan 0 artinya tidak)
3	<i>High blood pressure</i>	Apakah pasien memiliki hipertensi (boolean: 1 artinya iya dan 0 artinya tidak)
4	<i>Creatine phosphokinase (CPK)</i>	Kadar enzim CPK dalam darah (mcg/L)
5	<i>Diabetes</i>	Apakah pasien memiliki penyakit diabetes (boolean: 1 = Ya dan 0 = Tidak)
6	<i>Ejection Fraction</i>	Persentase darah yang meninggalkan jantung dalam setiap kontraksi
7	<i>Platelets</i>	Kadar keping darah atau trombosit dalam darah (kilo platelets/mL)
8	<i>Sex</i>	Jenis Kelamin (boolean: 1 = <i>Man</i> atau Laki-Laki dan 0 = <i>Woman</i> atau Perempuan)
9	<i>Serum Creatinine</i>	Kadar serum kreatinin dalam darah (mg/dL)
10	<i>Serum Sodium</i>	Kadar serum sodium dalam darah (mEq/L)
11	<i>Smoking</i>	Apakah pasien merokok (boolean: 1 = Ya dan 0 = Tidak)
12	<i>Time</i>	Durasi pengawasan pasien / <i>follow-up period</i> (hari)
13	<i>Death Event</i>	Apakah pasien meninggal dalam masa pengawasan / <i>follow-up</i> (boolean: 1 = meninggal dan 0 = tidak meninggal)

Selanjutnya kita akan memisahkan ke-13 atribut tersebut menjadi nilai atribut dan menyisakan satu atribut sebagai nilai tujuan. Nilai tujuan sendiri kami ambil dari atribut *DEATH EVENT*, agar kami dapat memprediksi apakah saat pasien mengalami gagal jantung akan menyebabkan kematian.

### 3. Pemodelan dengan Decision Tree

Penelitian ini akan melakukan pemodelan *decision tree* dengan menggunakan algoritma ID3 atau *iterative dichotomiser 3*. Algoritma ID3 ini melakukan pencarian dengan rakus atau *greedy* pada setiap kemungkinan pohon keputusan (Hikmatulloh et al. 2019). ID3 ini mencari atribut mana yang akan menjadi *root* atau akarnya. Dari akar ini lah nanti nya kita dapat menemukan atribut mana yang paling fatal yang dapat menyebabkan kematian pada pasien gagal jantung.

Cara kerja ID3 dapat dijabarkan sebagai berikut (Munthe et al. 2018):

- 1) Mencari node akar.
- 2) Mencari atribut dengan nilai information gain terbesar.
- 3) Membuat node dari atribut tersebut.
- 4) Melakukan perulangan mencari information gain terbesar sampai setiap node hanya memberikan satu label kelas.

Entropi merupakan formula untuk mengukur atau menghitung jumlah dari informasi yang ada (Munthe et al. 2018). Entropi sendiri dapat diukur sebagai berikut:

$$Entropi(S) = \sum_{j=1}^k -p_j \log_2 p_j$$

Dimana:

- $S$  : Himpunan (dataset) kasus
- $k$  : Banyaknya partisi  $S$
- $p_j$  : Probabilitas yang di dapat dari iya dibagi total kasus.

Untuk nilai entropi( $S$ ) = 0, hal ini terjadi dikarenakan semua contoh pada  $S$  berada pada kelas yang sama. Sedangkan untuk entropi( $S$ ) = 1, hal ini akan terjadi pada jumlah contoh pada  $S$  yang positif dan negatif adalah sama. Lalu yang terakhir  $0 < \text{entropi}(S) < 1$ , hal ini akan mungkin terjadi saat jumlah contoh pada  $S$  yang negatif dan positif tidak sama.

Information gain didasari oleh Entropi. Information gain adalah perbedaan antara entropi class dan entropi bersyarat kelas dengan fitur yang dipilih (Tangirala, 2020).

Setelah kita mendapatkan nilai entropi, kita akan menggunakan information gain agar dapat melakukan pemilihan atribut. Formula akan seperti berikut:

$$Gain(A) = Entropi(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropi(S_i)$$

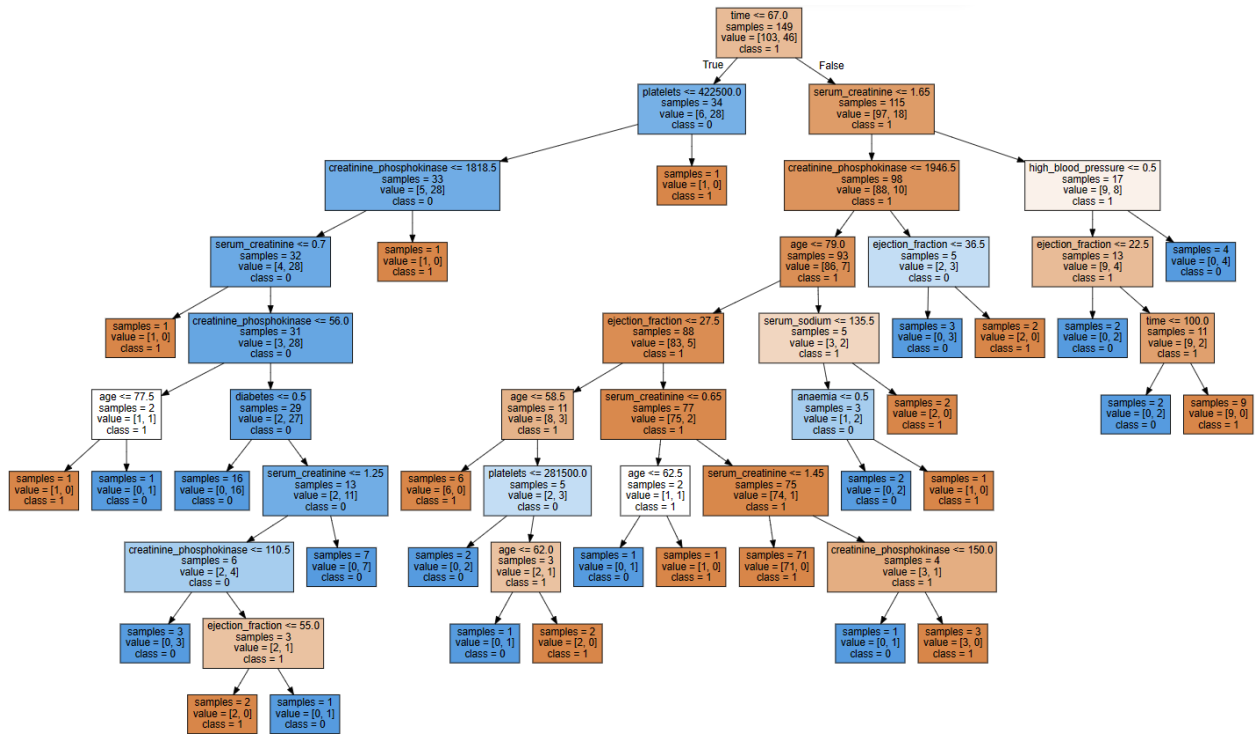
## Hasil dan Pembahasan

### 1. Pemrosesan Data

Dataset yang dimiliki diolah menggunakan collab researcher menghasilkan pohon keputusan untuk penentuan indikator letal gagal jantung dengan parameter penentunya berada pada Death Event dimana boolean 1 berarti pasien mengalami kematian dan sebaliknya untuk pengawasan dalam masa *follow-up*.

Dengan jumlah pasien 299 dalam dataset, kita akan membagi menjadi 2 bagian menggunakan split train. 50% atau sebanyak 149 data dari dataset akan digunakan untuk training. Lalu sisanya sebanyak 150 data akan digunakan untuk testing. Kita melakukan perhitungan *decision tree* dengan random state sebesar 500. Untuk nilai akurasinya sendiri didapatkan sebesar 0.76 atau 76%.

## 2. Hasil Pohon Keputusan



Gambar 2. Hasil Perhitungan Pohon Keputusan Menggunakan Collaboratory

Pohon keputusan atau *Decision Tree* ini menunjukkan indikator paling berpengaruh adalah time atau waktu *follow up* yang berada pada urutan atau *depth* paling atas pada pohon keputusan, dimana jika masa *follow-up* atau pengawasan  $\leq 67$  hari, prediksi atau pohon keputusan akan memasuki bagian kiri anak pohon keputusan dan sebaliknya untuk waktu pengawasan  $\geq 67$  hari maka prediksi atau hasil pohon keputusan akan memasuki bagian kanan anak pohon keputusan, sehingga jika sebuah indikator pada sebuah parameter benar, maka penentuan alur pohon keputusan akan otomatis berlanjut pada sisi kiri anak pohon keputusan dan jika sebuah indikator nya tidak sesuai dengan parameter nya maka penentuan alur pohon keputusan akan berlanjut ke sisi kanan anak pohon keputusan.

Selanjutnya, semakin bawahnya urutan baris atau *depth* sebuah parameter maka semakin rendah atau semakin tidak berpengaruhnya pula sebuah parameter dalam menandakan sebuah keletalan pada pasien dengan gagal jantung, sehingga, pada pohon keputusan yang dihasilkan di atas, parameter *ejection fraction* atau persentase darah

yang meninggalkan jantung dalam setiap kontraksi menjadi parameter yang paling tidak berpengaruh terhadap tingkat letal pada gagal jantung, dimana parameter tersebut berada pada urutan baris paling bawah atau *depth* ke-8 dengan total *sampel* atau kasus sebanyak 3 kasus yang terdiri dari 2 kasus untuk pasien yang meninggal dengan persentase *ejection fraction*  $\leq 55$  dan sisa 1 kasus untuk pasien tidak meninggal dengan persentase *ejection fraction*  $\geq 55$ .

Dua parameter yang memiliki pengaruh tertinggi dibawah dari *time* atau waktu *follow up*, adalah kadar keping darah (*platelets*) dan kadar serum kreatinin darah (serum creatinine) yang berada pada urutan atau *depth* kedua, dibawah parameter waktu (*time*). Sedangkan, untuk parameter *sex* / jenis kelamin dan *smoking* / status merokok pasien tidak menjadi parameter yang menentukan seberapa letal gagal jantung.

### **Kesimpulan**

Dari hasil pengklasifikasian menggunakan pohon keputusan (*decision tree*) diatas dapat disimpulkan bahwa dari semua parameter yang ada pada dataset yang diuji, parameter atau indikator paling berpengaruh terhadap tingkat letal penyakit gagal jantung adalah *time* atau waktu *follow-up* dimana parameter tersebut berada pada urutan atau *depth* paling atas dengan faktor *sex* dan *smoking* tidak menjadi parameter atau faktor yang letal pada penyakit gagal jantung. Di dapatkan juga bahwa tingkat akurasi pengklasifikasiannya sebesar 76%.

### **Ucapan Terima Kasih**

Puji syukur kepada Allah SWT yang telah memberikan rahmat dan karuniaNYA sehingga penulis dapat menyelesaikan jurnal ini. Tidak lupa kami juga mengucapkan banyak terima kasih kepada Bapak Dr. Wiharto, S.T., M.Kom. sebagai dosen mata kuliah kecerdasan buatan penulis atas bimbingan dan panduan selama proses pembuatan jurnal ini.



## Daftar Pustaka

- World Health Organization. 2021. Cardiovascular diseases (CVDs). URL: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) . Diakses tanggal 16 Juni 2022.
- Siswanto BB, Hersunarti N, Erwinanto, Barack R, Pratikto RS, Nauli SE, Lubis AC. 2015. Pedoman Tata Laksana Gagal Jantung: Perhimpunan Dokter Spesialis Kardiovaskular Indonesia. Jakarta: PP PERKI.
- Hermawati, F. Astuti. 2013. *Data Mining*. Edisi 1, Yogyakarta: Andi Offset.
- Hikmatulloh, H., Rahmawati, A., Wintana, D., & Ambarsari, D. A. 2019. Penerapan Algoritma iterative dichotomiser Three (ID3) Dalam mendiagnosa kesehatan kehamilan. KLIK - KUMPULAN JURNAL ILMU KOMPUTER, 6(2), 116. <https://doi.org/10.20527/klik.v6i2.189>
- Munthe, I. R., Sihombing, V. 2018. Klasifikasi Algoritma iterative Dichotomizer (ID3) Untuk Tingkat Kepuasan Pada Sarana laboratorium komputer. Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP), 1(2), 27–34. <https://doi.org/10.34012/jutikomp.v1i2.237>
- Tangirala, S. 2020. Evaluating the impact of gini index and information gain on classification using decision tree classifier algorithm\*. International Journal of Advanced Computer Science and Applications, 11(2). <https://doi.org/10.14569/ijacsa.2020.0110277>
- Suen, S.C., Chen, H.H., & Su, C.M. 2019. Predicting in-hospital mortality for heart failure patients using decision tree approach. Computer Methods and Programs in Biomedicine.
- Ridho, R., & Hendra, H. (2022). KLASIFIKASI DIAGNOSIS PENYAKIT COVID-19 MENGGUNAKAN METODE DECISION TREE. JUST IT : Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer, 11(3), 69–75. <https://doi.org/10.24853/JUSTIT.11.3>