ANALISIS SURVIVAL PASIEN *INSUFFICIENCIA CORDIS* MENGGUNAKAN MODEL REGRESI WEIBULL DAN MODEL REGRESI *COX PROPORTIONAL HAZARD*

D.Fidiana, *R. Budiarti, I G.P. Purnaba, dan N. Agustiani

Sekolah Sains Data, Matematika, dan Informatika,
Institut Pertanian Bogor, Jl. Meranti, Kampus IPB Dramaga Bogor.
dwifidiana@apps.ipb.ac.id, retnobu@apps.ipb.ac.id.***corresponding author**,
purnaba@apps.ipb.ac.id, nur_agustiani@apps.ipb.ac.id

Abstrak

Analisis *survival* digunakan untuk mengevaluasi durasi waktu dari awal individu masuk pengamatan hingga terjadinya suatu peristiwa, seperti kesembuhan atau kematian. Penelitian ini berfokus pada pasien *insufficiencia cordis*. Analisis dilakukan dengan pendekatan parametrik (regresi Weibull) serta semi-parametrik (*Cox Proportional Hazard*). Model regresi Weibull menjadi model terbaik dengan nilai AIC 127,50 dan MSE 0,5071. Variabel signifikan yang memengaruhi analisis *survival* pada penelitian ini adalah *age*, *ejection fraction*, *serum sodium*, *platelets*, dan *serum creatinine*. Penelitian ini memberikan kontribusi signifikan bagi dunia medis dan industri asuransi, memungkinkan identifikasi faktor risiko yang lebih akurat dan mendukung pengambilan keputusan dalam strategi penanganan medis serta penetapan premi asuransi yang berbasis risiko.

Kata kunci: Analisis *Survival*, *Cox Proportional Hazard*, *Insufficiencia Cordis*, Kaplan-Meier, Regresi Weibull.

1 Pendahuluan

Waktu merupakan elemen esensial dalam beragam aspek kehidupan, salah satunya pada bidang kesehatan. Dalam konteks statistik, waktu yang dihitung sejak awal pengamatan sampai suatu peristiwa tertentu terjadi dikenal sebagai waktu *survival* [1]. Metode statistika untuk mengevaluasi data dari variabel respon yang mengukur durasi waktu sejak awal pengamatan hingga terjadinya peristiwa, seperti penyembuhan, kambuhnya penyakit, atau kematian dinamakan analisis *survival*. Tujuannya adalah untuk mengestimasi peluang suatu peristiwa (seperti kesembuhan atau kematian) dalam periode waktu tertentu. Analisis ini juga bertujuan untuk mengevaluasi hubungan antara waktu terjadinya peristiwa (sebagai variabel respon) dengan faktor-faktor prediktor [2].

Salah satu kasus medis yang relevan untuk dianalisis dengan pendekatan ini adalah penyakit kardiovaskular, yang menjadi penyebab kematian utama di dunia [3]. Penyakit ini melibatkan penyakit lain seperti, jantung koroner, *insufficiencia cordis* (gagal jantung), hipertensi, dan *stroke*. Di antara berbagai bentuk penyakit kardiovaskular, *insufficiencia cordis* menjadi perhatian serius karena tingkat kematian yang tinggi dan kecenderungan peningkatan jumlah kasus dari tahun ke tahun. Faktor-faktor seperti kesehatan, jenis kelamin, usia, dan penyakit bawaan seperti diabetes juga memengaruhi risiko kematian akibat *insufficiencia cordis* [3].

2020 Mathematics Subject Classification: 62N02, 62P10.

Diajukan:12/06/2025, diterima: 30/06/2025. DOI: https://doi.org/10.29244/milang.21.1.45-59

Pendekatan model yang sering digunakan pada analisis *survival* diantaranya regresi Weibull dan *Cox Proportional Hazard* (CPH). Regresi Weibull adalah model parametrik yang berasumsi bahwa data mengikuti distribusi Weibull, untuk analisis data kegagalan material [2]. Sementara itu, model regresi CPH tidak mengasumsikan distribusi waktu *survival*, berbeda dengan model parametrik yang mengasumsikan distribusi tertentu [5]. Kedua Model regresi tersebut telah menjadi standar dalam berbagai bidang termasuk epidemiologi, onkologi, dan penelitian klinis, karena kemampuannya untuk mengevaluasi efek dari beberapa faktor risiko pada waktu *survival*.

Beberapa studi terdahulu telah membandingkan performa kedua model tersebut dalam konteks penyakit yang berbeda. Misalnya pada penelitian mengenai laju kesembuhan tuberkulosis paru, model regresi Weibull menunjukkan kinerja yang lebih baik dengan nilai *Akaike Information Criterion* (AIC) sebesar 133,828, lebih kecil dari nilai AIC CPH yang mencapai 325,809 [6]. Sebaliknya, dalam penelitian mengenai laju kesembuhan pasien demam berdarah *dengue*, CPH terbukti lebih baik dengan nilai AIC sebesar 134,932, sedangkan model regresi Weibull memiliki nilai AIC sebesar 177,245. Pemilihan model analisis *survival* harus disesuaikan dengan karakteristik data dan konteks penelitian untuk mencapai hasil yang optimal [5].

Penelitian ini bertujuan untuk menganalisis *survival* pasien *insufficiencia cordis* dengan menggunakan dua pendekatan, yaitu model regresi Weibull dan model regresi CPH. Hasil penelitian ini diharapkan dapat memberikan informasi model yang lebih sesuai untuk memprediksi *survival* pasien *insufficiencia cordis* berdasarkan karakteristik datanya. Informasi ini dapat dimanfaatkan oleh industri asuransi kesehatan dalam merancang premi berdasarkan risiko kematian, serta membantu dalam pengambilan keputusan terkait pengelolaan risiko dan biaya perawatan jangka panjang.

2 Metodologi

2.1 Analisis Survival

Terdapat tiga elemen utama yang penting dalam analisis *survival* yaitu waktu awal (*time origin*), waktu kegagalan (*failure time*), dan skala waktu [8]. Dalam metode analisis *survival*, terdapat fungsi utama yaitu fungsi *survival* dan fungsi *hazard*. Fungsi *survival* adalah fungsi yang menunjukkan kemungkinan waktu kegagalan lebih dari waktu t [3], yang dinyatakan sebagai berikut:

$$S(t) = 1 - F(t) \tag{1}$$

dengan $F(t)$ adalah fungsi sebaran yang menunjukkan peluang bahwa kegagalan terjadi sebelum waktu t [3].

Fungsi *hazard* menunjukkan peluang bahwa seseorang yang diasumsikan telah bertahan hingga waktu t akan mengalami kejadian pada waktu t sampai Δt[10]. Fungsi ini didefinisikan sebagai berikut:

$$\begin{aligned} h(t) &= \lim_{\Delta t} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = \lim_{\Delta t} \frac{F(t + \Delta t) - F(t)}{\Delta t} . \frac{1}{1 - F(t)} \\ &= \frac{F'(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \end{aligned} \tag{2}$$

dengan $f(t)$ adalah fungsi kepekatan peluang dari suatu peubah acak T yang didefinisikan sebagai limit dari peluang individu mengalami suatu kejadian tertentu dalam interval t sampai $t + \Delta t$ [19].

2.2 Kaplan-Meier dan Uji Log-Rank

Kaplan-Meier digunakan untuk mengestimasi fungsi *survival* dan fungsi *hazard* pada data waktu *survival* yang mungkin tersensor, dengan estimator sebagai berikut [11]:

$$\hat{S}(t_D) = \prod_{j=1}^{D} \frac{R(x_j) - d_j}{R(x_j)}, \tag{3}$$

dengan d_j adalah banyaknya kejadian pada waktu x_j, dan $R(x_j)$ adalah banyaknya individu yang masih dalam risiko (belum mengalami kejadian) tepat sebelum waktu x_j, dan t_D adalah waktu tertentu saat kejadian ke-D.

Uji *log-rank* merupakan metode non-parametrik yang didasarkan pada prinsip membandingkan jumlah kejadian pada setiap titik waktu di mana peristiwa terjadi [10]. Statistik uji *log-rank* dapat dihitung menggunakan rumus sebagai berikut:

$$\chi^2_{r,q} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{\sigma_i^2}, \tag{4}$$

dengan O_i adalah banyaknya kejadian yang diobservasi dalam kelompok i, E_i adalah banyaknya kejadian yang diharapkan dalam kelompok i, dan σ_i^2 merupakan ragam dari $O_i - E_i$. Hipotesis pada uji *log-rank* yaitu:

H_0: Tidak terdapat perbedaan dalam fungsi *survival* antara kelompok yang dibandingkan.

H_1: Terdapat perbedaan dalam fungsi *survival* antara minimal dua kelompok yang dibandingkan.

Menolak H_0 jika $\chi^2_{r,q} > \chi^2_{\text{tabel}}$ atau *p-value* $< \alpha$ yang ditetapkan. Notasi r merupakan banyaknya kelompok, dan q merupakan derajat bebas $(r - 1)$.

Pada penelitian ini, analisis dilakukan dengan pendekatan parametrik (regresi Weibull) serta pendekatan semi-parametrik (*Cox Proportional Hazard*). Regresi Weibull akan lebih baik dibandingkan regresi CPH jika terpenuhi asumsi bahwa waktu *survival* memiliki sebaran Weibull, sedangkan regresi CPH tidak membutuhkan asumsi sebaran.

2.3 Regresi *Cox Proportional Hazard*

Penelitian ini menggunakan pengujian asumsi melalui *Goodnesss of Fit* (GoF). Jika asumsi terpenuhi, maka pemodelan *Cox Proportional Hazard* (CPH) dan regresi Weibull dapat dilakukan. Pengecekan asumsi dilakukan dengan membentuk hipotesis dari uji *global test* sebagai berikut:

H_0: Data tidak memenuhi asumsi *proportional hazard*

H_1: Data memenuhi asumsi *proportional hazard*

Menolak H_0 pada asumsi *proportional hazard* terjadi jika *p-value* $> \alpha$ (taraf signifikansi).

Bentuk umum model CPH dengan *n* individu dan p kovariat yaitu sebagai berikut:

$$h(t) = h_0(t) \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi})\,;\ t \geq 0; i = 1,2, \ldots, n, \tag{5}$$

dimana, t adalah waktu, β_p merupakan parameter dari kovariat ke-p yang akan diestimasi dengan metode MLE, dan $h_0(t)$ adalah fungsi *hazard* dasar [6], dan $x_{pi} = [x_{1i}, x_{2i}, \ldots, x_{pi}]$ merupakan vektor data penelitian variabel kovariat ke-p pada suatu individu ke-i .

Fungsi kemungkinan untuk semua waktu kejadian adalah sebagai berikut:

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp(\sum_{k=1}^{p} \beta_k X_k)}{\sum_{m \in R_i} \exp(\sum_{k=1}^{p} \beta_k X_{(m)k})}, \tag{6}$$

dengan n adalah waktu kejadian yang berbeda, $X_{(m)k}$ adalah kovariat ke-k dari individu dengan waktu kejadian t_i, dan R_i adalah himpunan semua individu sebelum waktu kejadian i. Estimasi fungsi hazard untuk individu ke-i adalah sebagai berikut:

$$\widehat{h_i}(t,x) = \hat{h}_0(t) \exp(\hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_p x_{pi}) \ , t \geq 0; i = 1,2, \ldots, n, \tag{7}$$

dengan $\widehat{h_i}(t)$ adalah fungsi *hazard* dari estimasi suatu amatan $h(t)$ yang mengalami kejadian pada waktu t_i. Estimasi fungsi *hazard* dasar adalah sebagai berikut

$$\hat{h}_0(t_{(j)}) = 1 - \hat{\xi}_j, \tag{8}$$

perhitungan $\hat{\xi}_j$ dapat dilakukan menggunakan rumus berikut

$$\hat{\xi}_j = \left(1 - \frac{\exp(\hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_p x_p)}{\sum_{i \in R(t_j)} \exp(\hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_p x_{pi})}\right); \ i,j = 1,2, \ldots, n, \tag{9}$$

dengan $x_{pi} = [x_{1i}, x_{2i}, \ldots, x_{pi}]$ adalah vektor data penelitian variabel kovariat ke-p pada suatu individu ke-i yang mengalami kejadian (kematian, kegagalan, dll) pada saat waktu t_j. Selanjutnya, nilai-nilai $\hat{\beta}_i, i = 1, \ldots, p$ yang dihasilkan dengan memaksimalkan fungsi *log likelihood* dengan metode numerik [16].

Nilai *hazard ratio* digunakan untuk mengetahui resiko laju kematian pada suatu parameter [14]. *Hazard Ratio* dibagi menjadi dua bagian yaitu:

1. *Hazard Ratio* data kontinu

$$\widehat{HR} = e^{c\hat{\beta}} \tag{10}$$

2. *Hazard Ratio* data kategorik

$$\widehat{HR} = e^{\hat{\beta}} \tag{11}$$

dengan β adalah koefisien dari variabel prediktor yang menunjukkan perubahan *log-hazard* untuk setiap peningkatan satu unit pada suatu variabel, notasi c merupakan konstanta skala perubahan unit. *Hazard ratio* bernilai di bawah 1,0 memiliki dampak pada laju penurunan kematian, sama dengan 1,0 menunjukkan suatu variabel dari model tidak mempunyai dampak pada laju kematian di waktu t, di atas 1,0 menunjukkan kemungkinan variabel pada laju peningkatan kematian [14].

2.4 Regresi Weibull

Distribusi Weibull dengan dua parameter (λ, γ) [2], yaitu parameter skala (λ) dan parameter bentuk (γ), memiliki fungsi kepekatan peluang sebagai berikut:

$$f(t|\lambda,\gamma) = \frac{\gamma}{\lambda}\left(\frac{t}{\lambda}\right)^{\gamma-1} \exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}\right), \qquad t > 0 \tag{12}$$

Fungsi kumulatif distribusi Weibull dua parameter (λ, γ) adalah sebagai berikut:

$$F(t|\lambda,\gamma) \ = 1 - \exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}\right), \qquad t > 0 \tag{13}$$

Adapun fungsi *survival*nya yaitu:

$$S(t|\lambda,\gamma) = \exp\left(-\left(\frac{t}{\lambda}\right)^{\gamma}\right), \qquad t > 0 \tag{14}$$

Sedangkan untuk estimasi fungsi *hazard* distribusi Weibull:

$$h(t|\lambda,\gamma) = \lambda\gamma t^{\gamma-1}, \qquad t > 0 \tag{15}$$

Uji *Kolmogorov-Smirnov* digunakan untuk menguji distribusi sebaran data yang sesuai pada waktu *survival*. Uji *Kolmogorov-Smirnov* memiliki persamaan berikut [16].

$$D = \max_{x \le t \le u} |F_n(t) - F(t)| \tag{16}$$

Pengecekan asumsi dilakukan dengan membentuk hipotesis dari uji *Kolmogorov-Smirnov*. Hipotesis yang digunakan yaitu:

H_0: Waktu *survival* mengikuti distribusi Weibull

H_1: waktu *survival* tidak mengikuti distribusi Weibull.

Menolak H_0 jika $D_{hitung} > D_{tabel}$ atau *p-value* $< \alpha$. Statistik dari data yang tersedia pada penelitian ini, dinotasikan t adalah titik pemotongan, $x = 0$ jika data tidak terpotong, serta $u = \infty$ jika data tidak tersensor. $F_n(t)$ adalah nilai empiris distribusi kumulatif sampel dari waktu *survival*. $F(t)$ adalah fungsi distribusi kumulatif yang diuji.

Multikolineritas diuji menggunakan *Variance Inflation Factor* (VIF). Jika nilai VIF melebihi 10, maka terdapat multikolineritas [19]. Rumus VIF yaitu:

$$VIF_j = \frac{1}{1 - R_j^2} \tag{17}$$

dengan $R_j^2, j = 1,2 \dots, p$ adalah koefisien determinasi antara variabel ke-j dengan variabel bebas lainnya pada persamaan atau model.

Model regresi *survival* Weibull dan regresi *hazard* Weibull dapat dibentuk dengan beberapa kovariat dengan persamaan berikut [2]:

$$S(t,x) = \exp\left(-t^{\gamma} \exp\left(-\gamma\left(\beta_0 + \beta_1 x_{1\mathrm{i}} + \dots + \beta_p x_{pi}\right)\right)\right);\ i = 1,2,\dots,n, \tag{18}$$

$$h(t,x) = \lambda\gamma t^{\gamma-1} \exp\left(-\gamma\left(\beta_0 + \beta_1 X_{1\mathrm{i}} + \dots + \beta_p X_{pi}\right)\right);\ i = 1,2,\dots,n, \tag{19}$$

dengan $\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} t_i^{\gamma}\right)^{-\frac{1}{\gamma}}$ dan hasil estimasi parameter lainnya dilakukan dengan MLE.

Pendugaan nilai parameter β dapat diestimasi menggunakan fungsi *likelihood*. *Log-likelihood* pada regresi Weibull dapat dirumuskan sebagai berikut [2]:

$$l(\beta) = \sum_{i=1}^{n} \delta_i \left[-x_i^T\beta + \log\gamma + (\gamma - 1)\log t_i\right] + x_i^T\beta \log t_i^{\gamma}, \tag{20}$$

dengan δ_i pada formula *log-likelihood* adalah indikator status kejadian individu, dimana $\delta_i = 1$ menunjukkan kejadian (kematian) terjadi dan $\delta_i = 0$ menunjukkan data tersensor. Notasi β merupakan vektor prediktor linier dari $\beta_1, \beta_2, \ldots, \beta_p$. Parameter γ adalah parameter *shape* sebaran regresi Weibull. $x_i^T\beta$ adalah vektor kovariat individu ke-*i,* yang dikalikan dengan β untuk menghitung prediktor linier $(\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi})$, dengan p adalah kovariat variabel ke-p.

2.5 Uji Statistik Model

Setelah model terbentuk, dilakukan pengujian signifikansi parameter untuk mengetahui variabel yang berpengaruh signifikan terhadap waktu *survival.* Statistik uji yang digunakan yaitu uji rasio *likelihood*. Persamaan statistik uji yang digunakan yaitu:

$$G^2 = -2\ln\frac{L(\hat{\beta})}{L(\hat{\beta}_0)} \tag{21}$$

dengan $G \sim \chi_p^2$ dengan $\hat{\beta} = \left[\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p\right]^T$ dan $\hat{\beta}_0$ adalah *intercept* dari suatu model tersebut. Hipotesis dari uji rasio *likelihood* terhadap suatu parameter β_p yaitu:

H_0: $\beta_1 = \beta_2 = \ldots = \beta_p = 0$

H_1: minimal ada satu $\beta_k \neq 0, k = 1,2, \ldots, p$.

Menolak H_0 jika *p-value* $< \alpha$ (taraf signifikansi) atau $G_{hitung}^2 > \chi_{df,\alpha}^2$[2].

Menguji masing-masing parameter secara terpisah dan secara asimtotik menyebar normal baku dapat menggunakan uji Wald [10]. Statistik uji tersebut menyebar *chi-square* $(p \approx 1)$. Persamaan statistik uji wald yang digunakan yaitu:

$$W_j = \left[\frac{\widehat{\beta_J}}{SE(\widehat{\beta_J})}\right]^2, \tag{22}$$

dengan $SE(\widehat{\beta_J})$ adalah simpangan baku dari suatu penduga maksimum *likelihood* β_j, dimana $j = 1,2 \ldots, p$. Hipotesis dari Uji Wald terhadap suatu parameter $\widehat{\beta_J}$ yaitu:

H_0: $\widehat{\beta_J} = 0$

H_1: $\widehat{\beta_J} \neq 0$.

Menolak H_0 jika $W_j > W_{tabel}$ atau *p-value* $< \alpha$ (taraf signifikansi).

2.6 Seleksi Model

Model terbaik yang dipilih adalah model dengan nilai Akaike Information Criterion (AIC) terendah. Nilai AIC dirumuskan sebagai berikut:

$$AIC = -2\log\hat{L} + 2p, \tag{23}$$

dengan $\hat{L}$ adalah nilai *likelihood* dan p adalah banyaknya parameter.

Metrik evaluasi dalam analisis *survival* penelitian ini adalah *Mean Square Eror* (MSE), untuk mengukur kuadrat dari rata-rata yang merupakan selisih antara nilai yang diprediksi dengan nilai aktual, dengan persamaan sebagai berikut:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{24}$$

dengan y_i adalah data aktual ke-i, $\hat{y}_i$ estimasi data ke-i, dan n adalah banyaknya data.

3 Hasil dan Pembahasan

Penelitian ini menggunakan data sekunder yang diperoleh dari situs UCI *Machine Learning Repository*. Data merupakan rekam medis pasien penyakit *insufficiencia cordis* yang bersumber dari rumah sakit di Faisalabad, Pakistan yang diamati selama bulan April 2015 hingga Juli tahun 2016 dengan lama waktu 495 hari dan banyaknya pasien adalah 299 orang. Terdapat 13 variabel yang terdapat dalam dataset tersebut. Pada penelitian ini semua variabel digunakan untuk menganalisis faktor-faktor yang memengaruhi kelangsungan hidup pasien pada model Regresi Weibull dan *Cox Proportional Hazard* (CPH).

Tabel 1 Daftar variabel dataset pasien *insufficiencia cordis.*

Notasi	Variabel	Deskripsi	Kategori
x_1	*Age*	Usia pasien (Tahun)	-
x_2	*Anaemia*	Penurunan hemoglobin (boolean)	Pengidap anemia (1), Tidak mengidap anemia (0)
x_3	*Creatine Phosphokinase*	Tingkat enzim CPK dalam darah (mcg/L)	-
x_4	*Diabetes*	Penyakit bawaan *Diabetes* pasien (boolean)	Pengidap *Diabetes* (1), Tidak mengidap *Diabetes* (0)
x_5	*Ejection Fraction*	Persentase darah yang keluar dari jantung pada setiap kontraksi (persentase)	-
x_6	*High Blood Pressure*	Penyakit bawaan Hipertensi pasien (boolean)	Pengidap Hipertensi (1), Tidak mengidap Hipertensi (0)
x_7	*Platelets*	Trombosit dalam darah (kiloplatelet/L)	-
x_8	*Sex*	Jenis Kelamin (boolean)	Perempuan (0), Laki-laki (1)
x_9	*Serum Creatinine*	Tingkat kreatinin serum dalam darah (mg/dL)	-
x_{10}	*Serum Sodium*	Tingkat sodium serum dalam darah (mg/dL)	-
x_{11}	*Smoking*	Pasien merupakan seorang perokok/tidak	Perokok (1), Tidak Perokok
T_1	*Time*	Periode dirawat di rumah sakit (hari)	-
T_2	*Death Event*	*Event* pasien selama perawatan (boolean)	Pasien meninggal (1), pasien tersensor (0)

Tabel 1 menjelaskan 13 variabel dataset pasien *insufficiencia cordis*. Sebanyak 11 variabel merupakan variabel prediktor, dua lainnya merupakan variabel respon.

Statistik deskriptif digunakan untuk menggambarkan suatu karakteristik dasar seperti nilai maksimum, minimum, dan *mean* dari faktor-faktor yang diduga memengaruhi penyakit *insufficiencia cordis.* Pada Tabel 2 disajikan statistik deskriptif variabel prediktor dari dataset rekam medis 299 pasien *insufficiencia cordis*.

Tabel 2 Analisis deskriptif variabel prediktor kontinu.

Variabel	Minimum	Maximum	*Mean*
Age	40	95	60,83
Creatine Phosphokinase	23	786	581,8
Ejection Fraction	14	80	38,08
Platelets	0,025	0,850	0,263
Serum Creatinine	0,050	9,040	0,996
Serum Sodium	113	148	136,6
Time	2	487	127,8

Analisis deskriptif variabel kategorik pasien *insufficiencia cordis* disajikan pada Tabel 3 berikut.

Tabel 3 Analisis deskriptif variabel prediktor kategorik.

Variabel	Banyaknya pasien pada kategori (boolean)	
	0	1
Anaemia	170	129
Diabetes	174	125
High Blood Pressure	194	105
Sex	105	194
Smoking	203	96
Death Event	199	100

Tabel 3 menunjukkan bahwa terdapat 100 dari 299 pasien penderita *insufficiencia cordis* meninggal (terjadi *event*). Dari 299 pasien, mayoritas tidak memiliki penyakit bawaan seperti *anaemia*, *diabetes*, *high blood pressure*, dan mayoritas bukan perokok.

3.1 Hasil uji Log-Rank terhadap model Kaplan-Meier

Analisis Kaplan-Meier pada penelitian ini berfungsi untuk memvisualisasikan kurva waktu *survival* dan *event* yang terjadi. Hasil uji *log-rank* yang memiliki *p-value* kurang dari taraf signifikansi 10% dapat diinterpretasikan bahwa variabel tersebut memengaruhi waktu *survival* dan *event* pada pasien *insufficiencia cordis*. Berdasarkan hasil uji *log-rank* terhadap model Kaplan-Meier terdapat 6 variabel yaitu *age* (x_1), *creatine phosphokinase* (x_3), *ejection fraction* (x_5), *platelets* (x_7), *serum creatinine* (x_9), *serum sodium* (x_{10})yang memengaruhi *survival* hidup pasien *insufficiencia cordis.*

3.2 Pemodelan Regresi *Cox* Proportional Hazard

Asumsi *Proportional Hazard* merupakan asumsi yang harus terpenuhi sebelum memodelkan regresi CPH. Pengujian asumsi pada penelitian ini menggunakan uji *global test* atau *Goodnesss of Fit* (GoF).

Tabel 4 Hasil nilai *p-value Goodness of Fit.*

Notasi	Variabel	χ^2	Parameter	*p-value*
x_1	*Age*	0,0255	1	0,87
x_2	*Anaemia*	0,2121	1	0,65
x_3	*Creatine Phosphokinase*	0,5500	1	0,46
x_4	*Diabetes*	0,0250	1	0,87
x_5	*Ejection Fraction*	0,7130	1	0,40
x_6	*High Blood Pressure*	1,4612	1	0,23
x_7	*Platelets*	0,0273	1	0,87
x_8	*Sex*	0,1901	1	0,66
x_9	*Serum Creatinine*	0,1897	1	0,66
x_{10}	*Serum Sodium*	1,1613	1	0,28
x_{11}	*Smoking*	0,0504	1	0,82
	Global	6,7208	11	0,82

Berdasarkan Tabel 4, hasil *Goodness of Fit* untuk masing-masing variabel menghasilkan *p-value* lebih dari taraf signifikansi 10%. Dapat disimpulkan bahwa pada penelitian ini terpenuhi asumsi *proportional hazard* untuk semua variabel bebas. Dilihat dari nilai *p-value* yang dihasilkan, variabel-variabel tersebut diduga berpengaruh terhadap laju kematian pasien penderita *insufficiencia cordis,* serta tidak terdapat korelasi antara residual *Schoenfield* antar variabel prediktor.

Setelah dilakukan pemodelan awal CPH dengan uji asumsi sebelumnya, disimpulkan bahwa model awal sesuai dengan landasan teori dan minimal ada satu variabel prediktor yang signifikan. Untuk mengetahui variabel prediktor yang berpengaruh signifikan terhadap respon, dilakukan uji parsial. Hasil pengujian parsial model CPH yang didapatkan pada Tabel 5 menunjukkan analisis *survival* CPH pada taraf signifikansi 10% terdapat empat variabel yang berpengaruh signifikan pada laju kematian pasien penderita *insufficiencia cordis*. Variabel *Age* (x_1) *Ejection Fraction* (x_5), *Platelets* (x_7), *Serum Creatinine* (x_9) menghasilkan *p-value* kurang dari taraf signifikansi 10%, sehingga dapat disimpulkan bahwa empat variabel tersebut yang signifikan mempengaruhi peubah respon pada model CPH. Hal tersebut dapat diartikan bahwa empat variabel tersebut dapat memberikan kontribusi terhadap laju kematian pasien penderita *insufficiencia cordis.* Pemilihan model terbaik pada model CPH dengan metode eliminasi *backward.*

Tabel 5 Hasil pengujian signifikansi parameter model *Cox Proportional Hazard.*

Notasi	Variabel	$\hat{\beta}$	*p-value*
x_1	*Age*	0,03100	0,00015
x_2	*Anaemia*	0,29000	0,17000
x_3	*Creatine Phosphokinase*	0,00035	0,70000
x_4	*Diabetes*	0,01900	0,93000
x_5	*Ejection Fraction*	-0,04100	0,00028
x_6	*High Blood Pressure*	0,22000	0,28000
x_7	*Platelets*	0,12000	0,04000
x_8	*Sex*	0,00087	0,30000
x_9	*Serum Creatinine*	-0,07800	0,00026
x_{10}	*Serum Sodium*	-0,09800	0,65000
x_{11}	*Smoking*	-0,04200	0,85000

Model terbaik CPH dengan nilai AIC terendah sebesar 864,06 pada penelitian ini dapat ditulis dengan formula sebagai berikut.

$$h(t,x) = h_0(t)\exp(0{,}031x_{1i} - 0{,}041x_{5i} + 0{,}12x_{7i} - 0{,}078x_{9i} - 0{,}042x_{11i}) \qquad (25)$$

Masing-masing variabel yang dibentuk merupakan variabel yang dikategorikan signifikan pada model CPH. Peubah x_{1i} menyatakan umur pasien individu ke-i. Kesesuaian model CPH dapat diestimasi menggunakan metode uji rasio *likelihood* dan uji Wald. model tersebut memiliki nilai MLE *Test* dan Wald *Test,* masing masing memiliki nilai statistik 53,17 dan 53,19 dengan *p-value* kurang dari taraf signifikansi 10%. Hasil tersebut dapat diartikan bahwa minimal ada salah satu peubah *fit* terhadap model. Hal tersebut menunjukkan bahwa model CPH secara keseluruhan dapat berkontribusi terhadap laju kematian pasien penderita *insufficiencia cordis*. Nilai *hazard ratio* pada model CPH ditunjukkan pada Tabel 6.

Tabel 6 Nilai *hazard ratio* model *Cox Proportional Hazard* terbaik.

Variabel	$\hat{\beta}$	*Hazard Ratio*
Age	0,031	1,0314
Ejection Fraction	-0,041	0,9598
Platelets	0,120	1,1274
Serum Creatinine	-0,078	0,9249
Smoking	-0,042	0,9588

Tabel 6 menunjukkan bahwa nilai *hazard ratio* dari variabel-variabel pada model terbaik digunakan untuk mengestimasi kemungkinan terjadinya suatu peristiwa, yang dapat diamati berdasarkan perbedaan antar individu. Estimasi parameter untuk peubah *age* menghasilkan nilai *hazard ratio* sebesar 1,0314. Hal tersebut dapat diinterpretasikan bahwa laju kematian pasien penderita *insufficiencia cordis* semakin bertambah usia maka laju kematiannya akan meningkat sebesar $3{,}14\%$. Estimasi parameter untuk variabel *Ejection Fraction* menghasilkan nilai *hazard ratio* sebesar $0{,}9598$. Hal tersebut menunjukkan bahwa laju kematian pasien penderita *insufficiencia cordis* dengan persentase darah yang keluar dari jantung pada setiap kontraksi diluar ambang batas normal sebesar $50\% - 70\%$, maka memiliki risiko laju kematian lebih rendah sebesar $4{,}02\%$. Estimasi variabel *Platelets* menghasilkan nilai *hazard ratio* sebesar $1{,}1274$, hal tersebut dapat diartikan bahwa laju kematian pasien penderita *insufficiencia cordis* dengan tingkat trombosit dalam darah di atas ambang batas normal sebesar $0{,}15 - 0{,}45$ kiloplatelet/L memiliki risiko laju kematian lebih tinggi sebesar 12,74%. Estimasi parameter untuk variabel *Serum Creatinine* menghasilkan nilai *hazard ratio* sebesar $0{,}9249$. Hasil tersebut dapat diartikan bahwa laju kematian pasien penderita *insufficiencia cordis* dengan tingkat *Serum Creatinine* dalam darah diluar ambang batas normal sebesar $0{,}5 - 1{,}2$ mg/dL menurunkan risiko laju kematian lebih rendah sebesar 7,51%. Estimasi parameter untuk variabel *Smoking* menghasilkan nilai *hazard ratio* $0{,}9588$, hal tersebut mengindikasi bahwa laju kematian seseorang yang tidak merokok menurun sebesar 4,12% dibandingkan seorang perokok.

3.3 Pemodelan Regresi Weibull

Pengujian distribusi peubah respon yaitu pada variabel *time* dilakukan untuk memenuhi salah satu asumsi regresi weibul, yaitu untuk mengetahui distribusi sebaran Weibull pada data waktu *survival.* Uji Distribusi data dilakukan menggunakan uji *Kolmogorov-Smirnov*. Uji statistik *Kolmogorov-Smirnov* pada uji sebaran distribusi Weibull menghasilkan *p-value* sebesar $0{,}9789$. Hasil tersebut menyimpulkan bahwa peubah respon pada data pasien *insufficiencia cordis* menyebar distribusi sebaran Weibull dengan dua parameter yaitu *shape* $\hat{\gamma} = 1{,}5272$ dan *scale* $\hat{\lambda} = 141{,}6347$.

Pengujian multikolinieritas digunakan untuk mengetahui variabel prediktor pada penelitian ini tidak saling berkorelasi. Hasil Nilai VIF pada variabel yang digunakan pada penelitian ini ditunjukkan pada Tabel 7.

Tabel 7 Nilai VIF pada masing-masing variabel

Notasi	Variabel	VIF
x_1	*Age*	1,115954
x_2	*Anaemia*	1,225632
x_3	*Creatine Phosphokinase*	1,076852
x_4	*Diabetes*	1,171086
x_5	*Ejection Fraction*	1,175202
x_6	*High Blood Pressure*	1,114731
x_7	*Platelets*	1,159959
x_8	*Sex*	1,513358
x_9	*Serum Creatinine*	1,329482
x_{10}	*Serum Sodium*	1,178330
x_{11}	*Smoking*	1,340388

Tabel 7 menunjukkan hasil dari nilai VIF pada masing-masing variabel. Masing-masing variabel menunjukkan tidak adanya multikolinieritas antar variabel prediktor karena memiliki nilai VIF kurang dari 10. Hal tersebut dapat disimpulkan bahwa untuk semua peubah prediktor pada penelitian ini dapat dimodelkan pada model regresi Weibull. Selanjutnya, mengestimasi parameter model regresi. Hasil uji parsial pada model regresi Weibull ditunjukkan pada Tabel 8.

Hasil pengujian parsial model regresi Weibull yang didapatkan pada Tabel 8 menunjukkan hasil estimasi parameter $\hat{\beta}$ dan nilai *p-value* untuk masing-masing 11 variabel prediktor yang dianalisis. Variabel dengan nilai *p-value* lebih kecil dari taraf signifikansi, mengidikasikan variabel tersebut memiliki pengaruh signifikan pada laju kematian pasien penderita *insufficiencia cordis.*

Tabel 8 Hasil pengujian signifikansi parameter model regresi Weibull.

Notasi	Variabel	$\hat{\beta}$	*p-value*
x_1	*Age*	-0,0193000	0,00025*
x_2	*Anaemia*	-0,1248000	0,34726
x_3	*Creatine Phosphokinase*	-0,0000400	0,40501
x_4	*Diabetes*	-0,0119000	0,92743
x_5	*Ejection Fraction*	0,0255000	0,00005*
x_6	*High Blood Pressure*	-0,0268000	0,87441
x_7	*Platelets*	0,2610600	0,09217*
x_8	*Sex*	-0,0000047	0,50963
x_9	*Serum Creatinine*	-0,0727000	0,06491*
x_{10}	*Serum Sodium*	0,0379500	0,00239*
x_{11}	*Smoking*	0,2122000	0,88694

*variabel berpengaruh secara signifikan pada $\alpha = 10\%$

Penelitian ini dilakukan untuk menentukan model terbaik dari regresi Weibull dilakukan dengan melihat nilai Akaike's Information Criterion (AIC) terkecil menggunakan seleksi metode *backward*. Hasil pemilihan model terbaik ditunjukkan pada Tabel 9.

Tabel 9 Nilai AIC masing-masing model regresi Weibull.

Kombinasi Variabel	AIC
Semua variabel	128,51
Tanpa variabel x_4	128,31
Tanpa variabel x_4, x_{11}	128,11
Tanpa variabel x_4, x_6, x_{11}	127,91
$x_1, x_2, x_3, x_5, x_7, x_9, x_{10}$	127,76
$x_1, x_5, x_7, x_9, x_{10}$	127,50*

*model terbaik pada model Regresi Weibull

Tabel 9 menunjukan bahwa nilai AIC untuk semua variabel pada model regresi Weibull sebesar 128,51. Pada model regresi Weibull memiliki nilai *intercept* sebesar 0,087, serta nilai parameter *scale* sebesar 0,0602 dan *shape* sebesar 1,7123, didapatkan model terbaik pada model regresi Weibull dengan formula sebagai berikut.

$$S(t,x) = \exp(-t^{1,7123}\, exp(0{,}087 - 0{,}0193x_{i1} - 0{,}0255x_{i5} + 0{,}2610x_{i7} - 0{,}0727x_{i9} + 0{,}0379x_{i10})) \tag{26}$$

$$h(t,x) = 0{,}1032t^{0,7123} \exp(-\,1{,}7123(0{,}087 - 0{,}0193x_{i1} - 0{,}0255x_{i5} + 0{,}2610x_{i7} - 0{,}0727x_{i9} + 0{,}0379x_{i10})) \tag{27}$$

Pengujian model regresi Weibull menggunakan metode uji rasio *likelihood* dan uji Wald memiliki *p-value* masing-masing sebesar 7,7e-08 dan 2,69e-11. Pemodelan regresi Weibull yang dimodelkan dapat diartikan *fit* (cocok) pada dataset pasien penderita *insufficiencia cordis*.

Hazard ratio merupakan nilai untuk mengetahui laju kematian pasien penderita *insufficiencia cordis* dari variabel-variabel yang digunakan pada penelitian ini. Variabel yang diestimasi nilai *hazard ratio* merupakan variabel prediktor yang memiliki nilai AIC terendah pada pemilihan model terbaik regresi Weibull. Nilai *hazard ratio* ditunjukkan pada Tabel 13.

Tabel 13 Nilai *hazard ratio* model regresi Weibull terbaik.

Variabel	$\hat{\beta}$	*Hazard Ratio*
Age	-0,0193	0,9809
Ejection Fraction	0,0255	1,0258
Platelets	0,2610	1,2983
Serum Creatinine	-0,0727	0,9299
Serum Sodium	0,0379	1,0387

Tabel 13 menunjukkan bahwa nilai *hazard ratio* dari variabel dengan model terbaik regresi Weibull yang digunakan untuk mengukur laju kematian yang dapat dilihat dari antar individu. Estimasi parameter untuk peubah *age* menghasilkan nilai *hazard ratio* sebesar 0,9809. Hal tersebut dapat diinterpretasikan bahwa laju kematian pasien penderita *insufficiencia cordis* yang bertambah usia lebih rendah sebesar 1,91% dari *event*.

4 Perbandingan Kinerja Analisis Model

Pada penelitian ini, pemodelan yang dibentuk yaitu model regresi Weibull dan CPH signifikan pada taraf signifikansi 10%.Pemilihan model terbaik di antara model terbaik pada masing-masing model dilakukan dengan membandingkan nilai AIC dan MSE yang terkecil pada model, seperti dapat dilihat pada Tabel 14 berikut.

Tabel 14 Nilai AIC dan MSE kedua model.

Metrik	Model *Cox* PH	Model Regresi Weibull
AIC	864,06	127,50
MSE	3,0195	0,5071

Tabel 14 menunjukkan bahwa nilai AIC dan MSE pada model regresi Weibull berada di bawah nilai yang dihasilkan oleh model CPH. Model regresi Weibull memiliki nilai AIC 127,50 dan nilai MSE 0,5071. Hal tersebut menunjukkan nilai MSE kurang dari 10 dapat diartikan bahwa prediksi model sangat baik dan data yang digunakan pada penelitian ini sesuai dengan model regresi Weibull yang digunakan.

Berdasarkan hasil analisis *survival* penderita pasien penderita *insufficiencia cordis* diperoleh kesimpulan bahwa data waktu lamanya rawat inap pasien yang digunakan pada penelitian ini berdistribusi Weibull. Hasil penelitian dengan pembentukan model semiparametrik yaitu CPH menghasilkan nilai AIC yang terendah, yaitu sebesar 864,06 dan MSE sebesar 3,0195. Model regresi Weibull memiliki performansi nilai AIC 127,50 dan nilai MSE 0,5071. Oleh karena itu, model regresi Weibull memiliki nilai AIC dan MSE terkecil dibandingkan model CPH.

Model regresi Weibull memberikan hasil bahwa faktor yang memengaruhi laju kematian dan laju kesembuhan pasien penderita *insufficiencia cordis* adalah variabel *age, ejection fraction, serum sodium, platelets,* dan *serum creatinine.* Variabel yang memiliki nilai *hazard ratio* lebih dari satu yaitu *ejection fraction, platelets*, dan *serum sodium*, hal tersebut dapat diartikan bahwa variabel tersebut lebih berisiko menyebabkan laju kematian dibandingkan variabel prediktor lainnya. Selain itu, variabel yang memiliki nilai *hazard* ratio kurang dari satu yaitu *age* dan *serum creatinine* tidak lebih berisiko menyebabkan laju kematian dibandingkan variabel prediktor lainnya.

5 Simpulan

Berdasarkan hasil analisis *survival* penderita pasien penderita *insufficiencia cordis* diperoleh kesimpulan bahwa data waktu lamanya rawat inap pasien yang digunakan pada penelitian ini berdistribusi Weibull. Hasil penelitian dengan pembentukan model semiparametrik yaitu *Cox Proportional Hazard* menghasilkan nilai AIC sebesar 864,06 dan MSE sebesar 3,0195. Model regresi Weibull memiliki performansi nilai AIC 127,50 dan nilai MSE 0,5071. Oleh karena itu, model regresi Weibull memiliki nilai AIC dan MSE lebih kecil dibandingkan model *Cox Proportional Hazard*, artinya regresi Weibull lebih cocok digunakan untuk memodelkan data penelitian ini.

Faktor yang memengaruhi laju kematian dan laju kesembuhan pasien penderita *insufficiencia cordis berdasarkan* hasil pemodelan regresi Weibull adalah variabel *age, ejection fraction, serum sodium, platelets,* dan *serum creatinine.* Variabel yang memiliki nilai *hazard ratio* lebih dari satu yaitu *ejection fraction, platelets*, dan *serum sodium*, hal tersebut dapat diartikan bahwa variabel tersebut lebih berisiko menyebabkan laju kematian dibandingkan variabel prediktor lainnya. Selain itu, variabel yang memiliki nilai *hazard* ratio kurang dari satu yaitu *age* dan *serum creatinine* tidak lebih berisiko menyebabkan laju kematian dibandingkan variabel prediktor lainnya

Daftar Pustaka

[1] Azman MM. 2019. Analisa perbandingan nilai akurasi moving average dan exponential smoothing untuk sistem peramalan pendapatan pada perusahaan xyz. *Jurnal Sistem dan Informatika*. 13(2):36–45.

[2] Collett D. 2015. *Modelling Survival Data in Medical Research. Ed ke-3*. New York: CRC Press.

[3] Dwiyugo PH, Nurul HM. 2022. Weibull regression model on classified continuous data (case study: bod water pollution indicator at mahakam watersheds in 2016). *Jurnal Eksponensial*. 13(2):123–130.

[4] Erna WL, Dwi B. 2020. Analisis faktor yang memengaruhi laju kesembuhan pasien demam berdarah dengue menggunakan regresi cox proportional hazard dan regresi Weibull: studi kasus di kabupaten bantul, yogyakarta. *Jurnal Statistika Industri dan Komputasi*. 05(2):26.

[5] Faisal AR, Bustan MN, Annas S. 2020. Analisis survival dengan pemodelan regresi cox proportional *hazard* menggunakan pendekatan bayesian (studi kasus: pasien rawat inap penderita demam tifoid di rsud haji makassar). *Journal of Statistik and Its application on Teaching and Research*. 2(2):62. http://doi.org/10.35580/variansiunm14629

[6] Fernandes AAR, Solimun. 2016. *Pemodelan Statistika pada Analisis Reliabilitas dan Survival*. Malang: Universitas Brawijaya Press.

[7] Hamdanah FH, Fitrianah D. 2021. Analisis performansi algoritma linear regression dengan generalized linear model untuk prediksi penjualan pada usaha mikro, kecil, dan menengah. *Jurnal Nasional Pendidikan Teknik Informatika*. 10(1):23. http://doi/org/10.23887/janapati.v10i1.31035

[8] Jumayanti, Sunaryo E, Wicaksana A. 2020. Kualitas hidup pasien dengan penyakit kardiovaskular di yogyakarta. *Jurnal Kesehatan*. 13(1):1–12.

[9] Mensah F, Nimoh V, Turkson A. 2021. Handling cencoring and cencored data in survival analysis: a standalone systematic literature review. *International Journal of Mathematics and Mathematical Science*. 2021(1):1–16. http://doi.org/10.1155/2021/9307475

[10] Mufidah AS, Purhadi. 2016. Analisis survival pada pasien demam berdarah dengue di RS haji surabaya menggunakan model regresi Weibull. *Jurnal Sains dan Seni ITS*. 5(2):2337–3520.

[11] Monica ASY, Purhadi. 2015. Analisis faktor yang memengaruhi laju kesembuhan pasien tuberkulosis paru di rsud dr. Soetomo tahun 2015 menggunakan regresi Weibull dan regresi cox proportional hazard. *Jurnal Sains dan Seni ITS*. 2(5):2337–3520.

[12] Nisa Z, Poerwanto B, S MF. 2023. Survival analysis with cox proportional hazard model for tuberculosis patients. *Jurnal Varian*. 7(1):77–86. http://doi.org/10.30812/varian.v7i1.2994

[13] Pham H. 2019. A new criterion for model selection. *MDPI Journals*. 7(1215):1–12. http://doi.org/10.3390/MATH7121215

[14] Putri F, Martadiputra B. 2024. Log-rank test as a continuation of the Kaplan-Meier method in survival analysis using r language (case study: d-penicillamine treatment on the probabillity of survival in primary biliary cirrhosis (pbc) patients at the mayo clinic). *Jurnal Matematika, Statistika, dan Komputasi*. 21(1):285–306. http://doi.org/10.20956/j.v21i1.36073

[15] Rinne H. 2009. *The Weibull Distribution Handbook*. Boca Raton (Florida): Chapman & Hall/CRC.

[16] Sarumpaet AN, Wuryandari T, Sudarno. 2024. Penerapan model Weibull proportional *hazard* dan regresi cox proportional hazard pada kondisi financial distress. *Jurnal Gaussian*. 13(2): 450–461. http://doi.org/10.15710/j.gauss.13.2.450-461

[17] Solehah A, Fatekurohman M. 2018. Analisis survival hidup pasien kanker paru menggunakan regresi Weibull. *Indonesian Journal of Applied Statistics*. 1(2):79–87. http://doi.org/10.13057/ijas.v1i2.25276

[18] Turkson AJ, Ayiah-Mensah F, Nimoh V. 2021. Handling censoring and censored data in survival analysis: a standalone systematic literature review. *International Journal of Mathematics and Mathematical Sciences*. 2021. http://doi.org/10.1155/2021/9307475

[19] Walpole RE, Myers RH, Myers SL, Ye K. 2012. *Probability and statistiks for engineers and scientists: Ed ke-9*. Boston: Pearson Education, Inc.