
Analisis Ketahanan Hidup Pasien Kanker Paru Menggunakan Regresi Weibull

Arivatus Solehah¹ and Mohamat Fatekurohman²

^{1,2} Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember

mfatekurohman.fmipa@unej.ac.id

Abstract. Lung cancer is one of the diseases which difficult to detect because of uneasy symptoms detection till it develops being the risky one. But, if the disease has been found, it can spread fast and cause death. According to the data of WHO, the type of cancer which causes the most of death is lung cancer which reaches 1,3 million death per year. Therefore, a survival analysis will be conducted to determine factors that affect the survival of lung cancer patient by using Weibull regression. The result shows some factors that significantly influence the survival of lung cancer patient are gender, erythrocyte, and general condition.

Keywords : lung cancer, survival analysis, Weibull regression

1. Pendahuluan

Kanker paru adalah penyakit yang ditandai dengan tidak terkendalinya pertumbuhan sel dalam jaringan paru, terutama sel-sel yang melapisi bagian pernapasan [1]. Kanker paru menduduki peringkat ke-3 di antara kanker yang paling sering ditemukan di beberapa rumah sakit di Indonesia [2]. Pasien penderita kanker paru memerlukan penanganan dan tindakan yang cepat dan terarah. Oleh karena itu, perlu diketahui faktor-faktor yang mempengaruhi waktu ketahanan hidup pasien kanker paru. Adapun dalam ilmu statistika faktor-faktor yang mempengaruhi waktu ketahanan hidup pasien kanker paru-paru dapat dikaji menggunakan analisis *survival*.

Analisis *survival* merupakan suatu metode statistika yang digunakan untuk menguji kemampuan suatu sampel percobaan hingga muncul suatu kejadian yang ditentukan oleh peneliti [3]. Pemodelan *survival* terbagi menjadi tiga, yaitu model parametrik, model semiparametrik, dan model nonparametrik. Pada model parametrik, pengaruh variabel bebas terhadap waktu ketahanan hidup dapat diketahui dengan metode regresi parametrik. Model regresi parametrik yang paling banyak digunakan adalah regresi Weibull karena distribusi Weibull bersifat fleksibel. Distribusi Weibull bersifat fleksibel diakibatkan oleh adanya *shape parameter* yang β menentukan perubahan bentuk dari kurva *hazard*.

Penelitian sebelumnya terkait regresi Weibull dan regresi *Cox proportional hazard* dalam kasus tuberkulosis paru diperoleh kesimpulan bahwa regresi Weibull lebih baik dari regresi *Cox proportional hazard* dengan nilai AIC regresi Weibull sebesar 133,828, sedangkan nilai AIC regresi *Cox proportional hazard* sebesar 325,809 [4]. Selain itu,

dilakukan juga penelitian terkait kanker paru dihasilkan bahwa variabel yang signifikan mempengaruhi ketahanan hidup pasien adalah jenis kelamin, eritrosit, LED, status terapi, dan keadaan umum [5].

Berdasarkan uraian tersebut, peneliti ingin mengidentifikasi faktor-faktor yang berpengaruh nyata terhadap laju kesembuhan pasien kanker paru di rumah sakit Paru Jember menggunakan regresi Weibull. Penelitian ini diharapkan dapat memudahkan pihak rumah sakit dalam menganalisis waktu ketahanan hidup pasien kanker paru di rumah sakit Paru Jember agar lebih memperhatikan faktor-faktor yang berpengaruh nyata dalam waktu ketahanan hidup pasien.

2. Metode Penelitian

Sebelumnya akan dijelaskan mengenai konsep dasar dalam analisis tahan hidup dan distribusi Weibull. Fungsi-fungsi pada distribusi waktu tahan hidup merupakan suatu fungsi yang menggunakan variabel random kontinu non negatif. Variabel random kontinu non negatif biasanya dinotasikan dengan huruf T dengan T dalam interval $[0, \infty)$. Fungsi yang saling ekuivalen tersebut adalah fungsi tahan hidup, fungsi kepadatan peluang, dan fungsi *hazard*.

Menurut Kleinbaum and Klein [6], fungsi *survival* $S(t)$ memberikan probabilitas bahwa seseorang bertahan lebih lama dari beberapa waktu yang ditentukan t dan dirumuskan sebagai

$$S(t) = 1 - F(t)$$

sedangkan $F(t)$ adalah fungsi distribusi kumulatif yaitu

$$F(t) = \int f(x)dx$$

Adapun $f(x)$ notasi dari fungsi kepadatan peluang yang didefinisikan sebagai peluang suatu individu mengalami kejadian (*event*), gagal atau mati dalam interval waktu t sampai $(t + \Delta t)$. $f(t)$ dirumuskan sebagai

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr[t \leq T < t + \Delta t]}{\Delta t}$$

Distribusi Weibull merupakan salah satu distribusi kontinu yang banyak digunakan dalam pemodelan analisis kelangsungan hidup. Misal T adalah waktu *survival* dan T berdistribusi Weibull, maka T merupakan suatu variabel random positif, dengan parameter $\gamma > 0$ dan $\beta > 0$, jika fungsi kepadatan peluangnya diberikan oleh

$$f(t) = \begin{cases} \frac{\gamma}{\beta} \left(\frac{t}{\beta}\right)^{\gamma-1} e^{-\left(\frac{t}{\beta}\right)^\gamma}, & t > 0 \\ 0, & \text{lainnya} \end{cases}$$

Akaike's Information Criterion (AIC) adalah metode yang berguna untuk mendapatkan model terbaik yang ditemukan oleh Akaike. Nilai AIC dapat dilihat pada persamaan berikut

$$AIC = -2l(\hat{\beta}) + 2df$$

dengan $l(\hat{\beta})$ adalah fungsi *log likelihood* dan df adalah total derajat bebas yang digunakan dalam model. Model regresi ataupun distribusi terbaik adalah model regresi yang memiliki nilai AIC terkecil. Model dengan nilai AIC yang terkecil merupakan model terbaik.

Data yang digunakan dalam penelitian ini adalah data sekunder yakni data pasien kanker paru di rumah sakit Paru Jember periode Januari 2016 hingga Mei 2016 dengan variabel responnya waktu ketahanan hidup pasien hingga meninggal. Variabel bebas yang digunakan ada 10 dengan masing-masing variabel telah dikategorikan, yaitu jenis kelamin terdiri dari kategori laki-laki dan perempuan; umur terdiri dari kategori $15 \leq \text{umur} < 25$, $25 \leq \text{umur} < 45$, $45 \leq \text{umur} < 65$, dan $\text{umur} \geq 65$; HB terdiri dari kategori kurang, normal, dan lebih; leukosit terdiri dari kategori kurang, normal, dan lebih; eritrosit terdiri dari kategori kurang, normal, dan lebih; LED terdiri dari kategori normal dan lebih; status terapi terdiri dari kategori ikut dan tidak ikut; riwayat penyakit terdiri dari kategori punya dan tidak punya; keadaan umum terdiri dari lemah, cukup, dan baik; dan BB terdiri dari kategori $BB < 30$, $30 \leq BB < 40$, $40 \leq BB < 50$, $50 \leq BB < 60$, dan $BB \geq 60$.

Tahap-tahap dalam menganalisis data pada penelitian ini yaitu pertama melakukan studi literatur; kedua pengumpulan data sekunder yang diperoleh dari penelitian sebelumnya; ketiga mendeskripsikan karakteristik pasien kanker paru berdasarkan waktu *survival* dan faktor yang diduga mempengaruhi ketahanan hidupnya; keempat menggambarkan kurva Kaplan-Meier pasien berdasarkan 10 variabel bebas; kelima melakukan uji *log-rank* untuk mengetahui adanya perbedaan antar kurva ketahanan hidup; keenam melakukan uji kesesuaian distribusi data melalui pendekatan *Kolmogorov-Smirnov* untuk mengetahui distribusi yang paling sesuai pada data waktu ketahanan hidup; ketujuh mendeteksi kasus multikolinieritas untuk mengetahui hubungan antarvariabel pada data penelitian; kedelapan mencari model terbaik menggunakan kriteria AIC pada regresi Weibull; kesembilan melakukan uji signifikansi parameter secara serentak dan parsial, akhirnya didapatkan model ketahanan hidup dari hasil estimasi parameter model terbaik regresi Weibull.

3. Hasil dan Pembahasan

3.1. Analisis Deskriptif

Data yang akan dianalisis adalah data pasien penderita kanker paru-paru di rumah sakit Paru Jember pada bulan Januari 2016 sampai Mei 2016 yang diperoleh dari penelitian sebelumnya. Adapun sensor yang digunakan pada penelitian ini adalah jenis penyensoran tipe III. Adanya data tersensor dikarenakan terdapat pasien menyatakan tersensor jika keluar dari rumah sakit dalam keadaan masih hidup, rawat jalan, ataupun dirujuk ke rumah sakit lain. Deskripsi variabel bebas yang diduga berpengaruh signifikan terhadap waktu ketahanan hidup pasien penderita kanker paru-paru dapat dijelaskan pada Tabel 1.

Berdasarkan Tabel 1 dapat diperoleh informasi bahwa rata-rata pasien kanker paru di rumah sakit Paru jember dapat bertahan hidup selama 6 hari. Selanjutnya, peluang pasien berjenis kelamin perempuan cenderung cepat meninggal dari pada pasien berjenis kelamin laki-laki yaitu sebesar 80,65%. Mayoritas pasien kanker paru yang meninggal berusia lebih atau sama dengan 25 atau kurang dari 45, yaitu sebesar 92,86%. Pasien kanker paru dengan HB kurang cenderung cepat meninggal yaitu sebesar 85,19%. Peluang meninggal 100% dimiliki oleh pasien dengan jumlah leukosit kurang, artinya pasien kanker paru yang kekurangan leukosit tidak dapat bertahan hidup. Pasien kanker paru dengan jumlah eritrosit kurang cenderung cepat meninggal yaitu sebesar 92,86%. Peluang meninggal 100% dimiliki oleh pasien dengan LED normal, artinya pasien kanker paru tidak dapat bertahan hidup dengan LED normal atau pasien kanker paru membutuhkan LED lebih untuk bertahan hidup. Pasien yang tidak ikut terapi memiliki peluang meninggal lebih besar yaitu 80%. Selain itu, pasien yang memiliki riwayat penyakit lain juga berkemungkinan lebih cepat meninggal dari pada pasien yang tidak memiliki riwayat penyakit lain yaitu sebesar 75,93%. Apabila keadaan umum pasien lemah, maka peluang meninggalnya lebih cepat, namun sebaliknya jika keadaan umum pasien tergolong baik maka pasien tetap hidup sampai akhir penelitian. Peluang meninggal 100% juga dimiliki bagi pasien dengan BB kurang dari 30 dan lebih atau sama dengan 60, artinya setiap pasien yang memiliki BB kurang dari 30 dan lebih atau sama dengan 60 dapat dipastikan meninggal dalam selang waktu penelitian yang telah ditentukan.

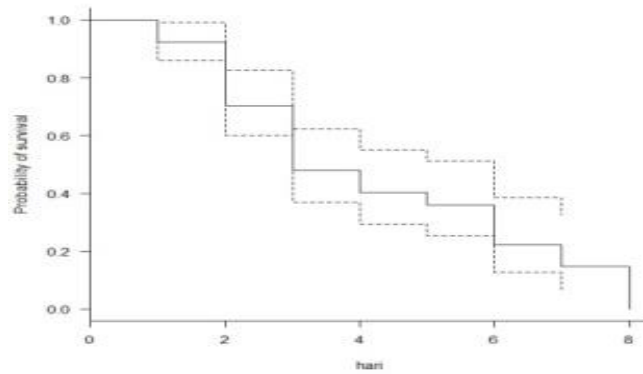
Tabel 1. Analisis Deskriptif

Variabel	Kategori	Jumlah pasien	Mean (hari)
Waktu ketahanan hidup	-	64	3,594
Jenis Kelamin	Perempuan	31	0,8065

	Laki-laki	33	0,697
Umur	15≤umur<25	9	0,7778
	25≤umur<45	14	0,9286
	45≤umur<65	29	0,5862
	umur≥65	12	0,9167
HB	Kurang	27	0,8519
	Normal	35	0,6857
	Lebih	2	0,50
Leukosit	Kurang	1	1
	Normal	24	0,7917
	Lebih	39	0,7179
Eritrosit	Kurang	14	0,9286
	Normal	19	0,7368
	Lebih	31	0,6774
LED	Normal	10	1
	Lebih	54	0,7037
Status terapi	Ikut terapi	59	0,7458
	Tidak terapi	5	0,8
Riwayat penyakit	Memiliki penyakit lain	54	0,7593
	Tidak memiliki penyakit lain	10	0,70
Keadaan umum	Kurang/lemah	24	0,9583
	Cukup	39	0,641
	Baik	1	0
BB	BB < 30	1	1
	30 ≤ BB < 40	21	0,8095
	40 ≤ BB < 50	32	0,6875
	50 ≤ BB < 60	8	0,75
	BB ≥ 60	2	1

3.2. Analisis Kaplan Meier dan Uji Log Rank

Analisis Kaplan-Meier digunakan untuk mengestimasi fungsi ketahanan hidup pasien kanker paru. Kurva ketahanan hidup pada Gambar 1 menginformasikan bahwa semakin besar waktu ketahanan hidup pasien penderita kanker paru, maka peluang pasien bertahan hidup cenderung semakin kecil. Hal tersebut berarti peluang pasien mengalami perbaikan kondisi klinis sampai waktu ke T semakin kecil mendekati nol.



Gambar 1. Kurva Kaplan Meier Waktu Tahan Hidup

Berdasarkan hasil uji *log-rank* pada Tabel 2 diperoleh bahwa variabel jenis kelamin, HB, leukosit, eritrosit, LED, status terapi, riwayat penyakit, dan BB tidak ada perbedaan antara kurva ketahanan hidupnya, sehingga dapat disimpulkan variabel jenis kelamin, HB, leukosit, eritrosit, LED, status terapi, riwayat penyakit, dan BB memiliki peluang ketahanan hidup cenderung sama untuk masing-masing kategori. Sementara itu, variabel umur dan keadaan umum terdapat perbedaan yang signifikan antara kurva ketahanan hidup, sehingga dapat disimpulkan bahwa variabel umur dan keadaan umum memiliki peluang ketahanan hidup yang berbeda.

Tabel 2. Statistik Uji *Log Rank* Sepuluh Variabel

Variabel	Statistik uji	χ^2_{tabel}	<i>p-value</i>
Jenis kelamin	1,8	3,841	0,176
Umur	12,5	7,815	0,00583
HB	0,5	5,991	0,767
Leukosit	0,3	5,991	0,849
Eritrosit	0,9	5,991	0,649
LED	3	3,841	0,0838
Status terapi	1,3	3,841	0,26
Riwayat penyakit	0,1	3,41	0,76
Kedaaan umum	10,3	5,991	0,00573
BB	4,4	9,488	0,349
BB	4,4	9,488	0,349

3.3. Uji Kesesuaian Distribusi Data

Pengujian kesesuaian distribusi data pada penelitian ini menggunakan *Kolmogorov-Smirnov*.

Weibull [#59]					
Kolmogorov-Smirnov					
Sample Size	64				
Statistic	0,18627				
P-Value	0,02039				
Rank	31				
α	0,2	0,1	0,05	0,02	0,01
Critical Value	0,13151	0,15027	0,16693	0,18667	0,20029
Reject?	Yes	Yes	Yes	No	No

Gambar 2. Uji *Kolmogorov-Smirnov* Data Waktu Tahan Hidup

Uji *Kolmogorov-Smirnov* pada Gambar 2 diperoleh bahwa $\alpha = 2,0785$, $\beta = 3,9995$, nilai $p - value = 0,02039$, $D_{hit} = 0,18627$. Jika $D_{tabel} = 0,189678$, maka berdasarkan hasil tersebut didapatkan bahwa $p - value > \alpha(0,01)$ atau $D_{hit} < D_{tabel}$, artinya H_0 diterima yaitu data waktu ketahanan hidup pasien penderita kanker paru-paru di rumah sakit Paru Jember mengikuti distribusi Weibull. Berdasarkan hasil uji kesesuaian, distribusi Weibull data waktu ketahanan hidup pasien kanker paru-paru dapat dilihat melalui histogram pada Gambar 2.

Tabel 3. p -value Uji *Chi-Square* Sepuluh Variabel

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
X_1	0	0,4173	0,1593	0,5681	0,2399	0,8128	0,315	0,8128	0,5681	0,0539
X_2	0,4173	0	0,4593	0,3468	0,1028	0,4373	0,3348	0,2097	0,0355	0,6597
X_3	0,1593	0,4593	0	0,7538	0,00105	0,7447	0,9032	0,3958	0,7497	0,5354
X_4	0,5681	0,3468	0,7538	0	0,5912	0,9018	0,6534	0,9018	0,5911	0,6971
X_5	0,2399	0,1028	0,00105	0,5912	0	0,0504	0,8479	0,3276	0,1563	0,7788
X_6	0,8128	0,4373	0,7447	0,9018	0,0504	0	0,7183	0	0,9018	0,7507
X_7	0,315	0,3348	0,9032	0,6534	0,8479	0,7183	0	0,3565	0,6534	0,8901
X_8	0,8128	0,2097	0,3958	0,9018	0,3276	0	0,3565	0	0,77	0,3125
X_9	0,5681	0,0355	0,7497	0,5911	0,1563	0,9018	0,6534	0,77	0	0,7239
X_{10}	0,0539	0,6597	0,5354	0,6971	0,7788	0,7507	0,8901	0,3125	0,7239	0

3.4. Pemilihan Model Terbaik

Kriteria yang digunakan dalam pemilihan model terbaik dilihat dari nilai AIC dari semua model hasil kombinasi variabel bebas. Model terbaik diperoleh dari model dengan nilai AIC terkecil sebesar 217,8235 yaitu

$$\hat{h}(t) = \frac{\gamma}{\Delta} t^{\gamma-1}$$

dengan

$$\Delta = [\exp(\alpha_0 + \alpha_1 X_{1(2)} - \alpha_5 X_5 + \alpha_9 X_9)]^{2,1186}.$$

3.5. Uji Signifikansi

Uji signifikansi parameter dalam model terbaik dilakukan secara serentak dan parsial. Pada uji serentak dihasilkan bahwa $G_{hitung}^2 = 14,6522$ dan $p - value = 0,002139163$, sedangkan $\chi_{3;0,01}^2$ pada tabel sebesar 11,34 sehingga diperoleh $G_{hitung}^2 > \chi_{3;0,01}^2$ atau $p - value < 0,01$ sehingga dapat disimpulkan bahwa variabel X_1, X_5, X_9 minimal ada satu yang berpengaruh dalam model. Berdasarkan hasil uji parsial Tabel 4 dapat diketahui bahwa variabel jenis kelamin, eritrosit, dan keadaan umum memiliki nilai-nilai $p - value$ kurang dari $\alpha(0,01)$ sehingga diputuskan tolak H_0 . Penolakan terhadap H_0 memberikan arti bahwa variabel tersebut berpengaruh terhadap model.

Tabel 4. Uji Parsial Model Terbaik

Variabel	Estimasi Parameter	Std. Error	Z	P-Value
Intercept (μ)	0,656	0,3546	1,85	$6,43 e^{-02}$
Jenis kelamin(2)	0,243	0,1385	1,75	$7,98 e^{-02}$
Eritrosit	-0,145	0,0841	-1,72	$8,46 e^{-02}$
Keadaan umum	0,509	0,103	3,63	$2,87 e^{-04}$
Scale (σ)	0,472			

Pada Tabel 4 ditunjukkan bahwa hasil estimasi dari *intercept* (μ) adalah 0,656 dan estimasi dari parameter *scale* (σ) adalah 0,472, sehingga dapat dihitung untuk parameter Weibull *shape* ($\hat{\beta}$) diperoleh

$$\hat{\gamma} = \frac{1}{\hat{\sigma}} = \frac{1}{0,472} = 2,1186$$

Jadi, model *hazard* dari regresi Weibull adalah

$$\hat{h}(t) = \frac{\gamma}{\beta^\gamma} t^{\gamma-1}$$

$$\hat{h}(t) = \frac{2,1186}{\Delta} t^{1,1186}$$

dengan

$$\Delta = [\exp(0,656 + 0,243X_{1(2)} - 0,145X_5 + 0,509X_9)]^{2,1186}.$$

4. Kesimpulan

Model regresi Weibull terbaik diperoleh dengan nilai AIC 217,8235 dan faktor-faktor yang berpengaruh secara signifikan terhadap waktu ketahanan hidup pasien penderita kanker paru-paru adalah jenis kelamin, eritrosit, dan keadaan umum.

Daftar Pustaka

- [1] Hashemi, A., Pilevar, A.H., and Rafeh, R. *Mass Detection in Lung CT Images Using Region Growing Segmentation and Decision Making Based on Fuzzy Inference System and Artificial Neural Network*. I.J. Image, Graphics and Signal Processing. Vol 6: 16-24. 2013.
- [2] Hulma, M.A., Basyar, M., dan Mulyani, H. Hubungan Karakteristik Penderita dengan Gambaran Sifatologi Pada Kasus Karsinoma Paru yang Dirawat di RSUD Dr. M. Jamil Padang. *Jurnal*. 3 (2): 196-201. 2014.
- [3] Collet, D. *Modelling Survival Data In Medical Research*. ed. London: Chapman and Hall. 2003.
- [4] Monica, A.S.Y. dan Purhadi. Analisis Faktor yang Mempengaruhi 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. 2016.
- [5] Fatekurohman, M., Nurmala, N., and Anggraini, D. Comparison of exact, efron and breslow parameter approach method on hazard ratio and stratified cox regression model. *Journal of Physics Conference Series*, 1008012007. doi: 10.1088/1742-6596/1008/1/012007. 2018.
- [6] Kleinbaum, D.G. and Klein, M. *Survival Analysis a Self-Learning Text Third Edition*. New York: Springer. 2012.