



**PERBANDINGAN METODE *COX PROPORTIONAL HAZARD*
DAN *RANDOM SURVIVAL FOREST* UNTUK MENGANALISIS
KETAHANAN HIDUP PASIEN PENDERITA KANKER
PAYUDARA**

SKRIPSI

Oleh

Ernanda Kusuma Widanaryanti

NIM 191810101089

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS JEMBER

2023



**PERBANDINGAN METODE *COX PROPORTIONAL HAZARD*
DAN *RANDOM SURVIVAL FOREST* UNTUK MENGANALISIS
KETAHANAN HIDUP PASIEN PENDERITA KANKER
PAYUDARA**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

Oleh

Ernanda Kusuma Widanaryanti

NIM 191810101089

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

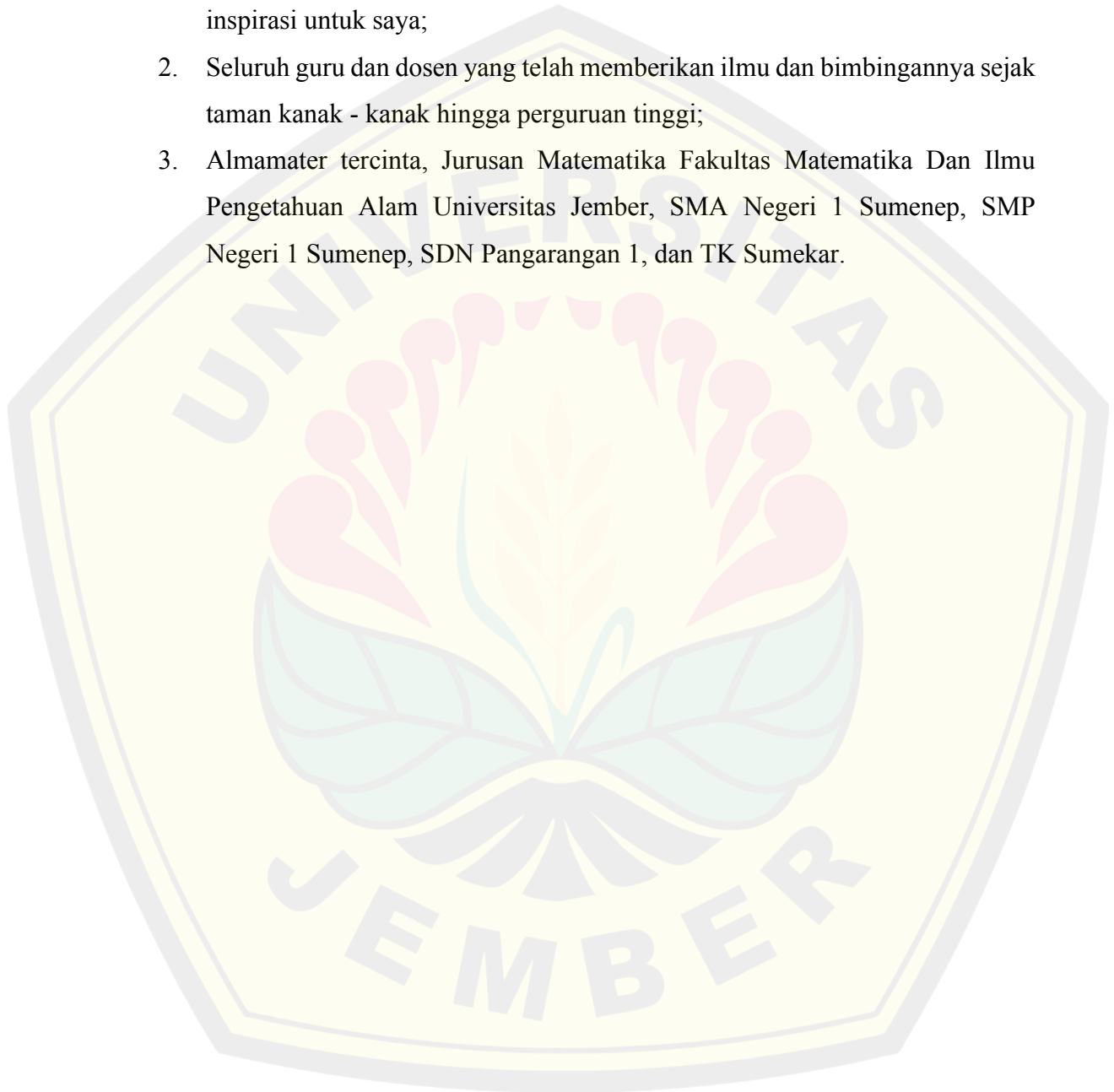
UNIVERSITAS JEMBER

2023

PERSEMBAHAN

Puji syukur kehadirat Tuhan Yang Maha Esa, skripsi ini saya persembahkan untuk:

1. Ayahanda Kusnadi, ibunda Ida Khotimah, adik tercinta Erinda Kusuma Larasati serta keluarga besar saya yang telah memberikan do'a, motivasi dan inspirasi untuk saya;
2. Seluruh guru dan dosen yang telah memberikan ilmu dan bimbingannya sejak taman kanak - kanak hingga perguruan tinggi;
3. Almamater tercinta, Jurusan Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Jember, SMA Negeri 1 Sumenep, SMP Negeri 1 Sumenep, SDN Pangarangan 1, dan TK Sumekar.



MOTTO

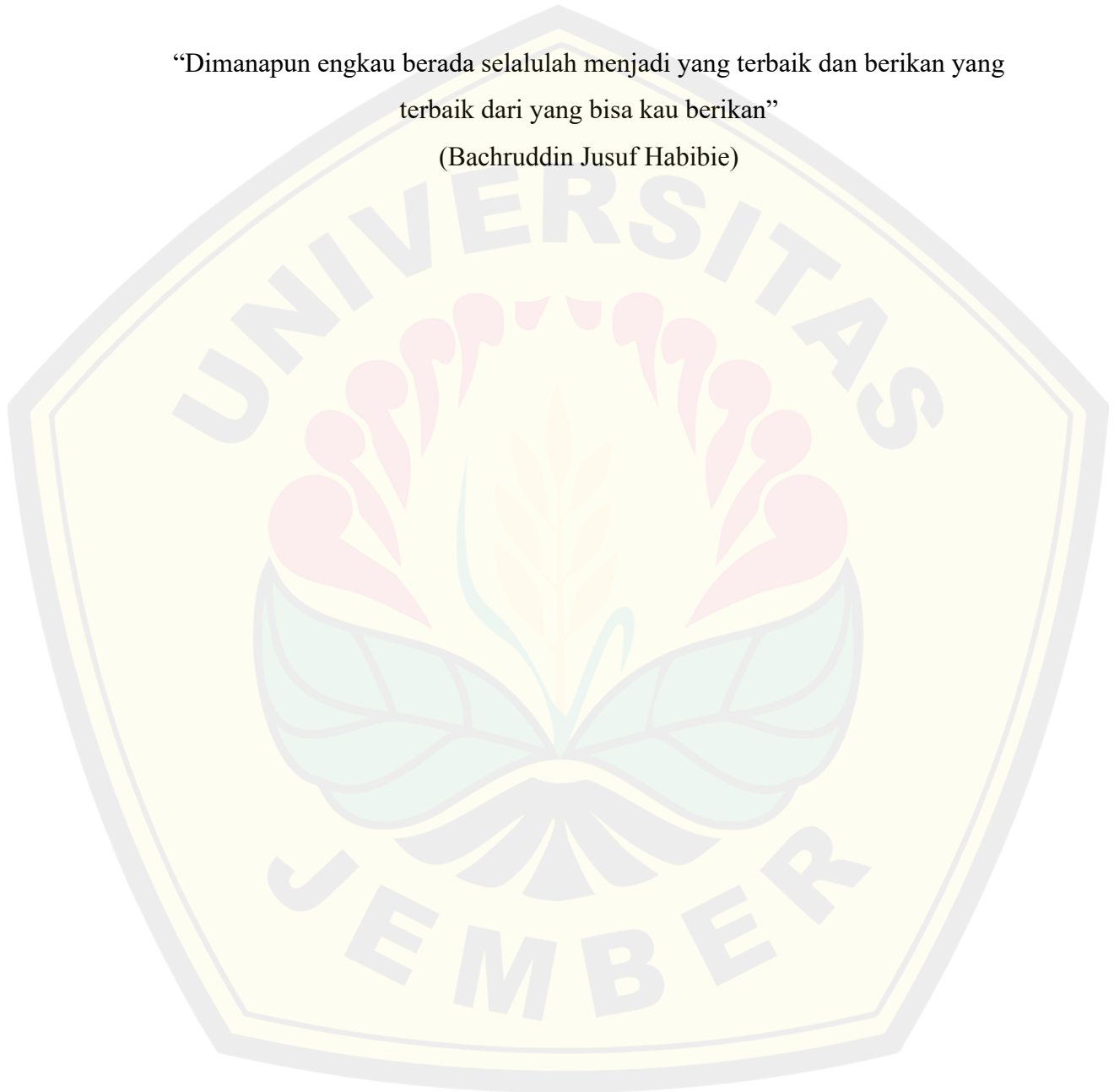
حَيْرُ النَّاسِ أَنْفَعُهُمْ لِلنَّاسِ

“Sebaik-Baik Manusia Adalah Yang Paling Bermanfaat Bagi Orang Lain”

(HR. Ahmad)

“Dimanapun engkau berada selalulah menjadi yang terbaik dan berikan yang terbaik dari yang bisa kau berikan”

(Bachruddin Jusuf Habibie)



PERNYATAAN

Saya yang bertandatangan di bawah ini:

Nama : Ernanda Kusuma Widanaryanti

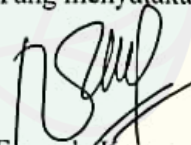
NIM : 191810101089

menyatakan dengan sesungguhnya bahwa karya tulis ilmiah yang berjudul “Perbandingan Metode *Cox Proportional Hazard* dan *Random Survival Forest* Untuk Menganalisis Ketahanan Hidup Pasien Penderita Kanker Payudara” adalah benar - benar hasil karya sendiri, kecuali jika disebut sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian surat pernyataan ini saya buat dengan sebenarnya, tanpa ada tekanan dan paksaan dari pihak mana pun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, April 2023

Yang menyatakan,

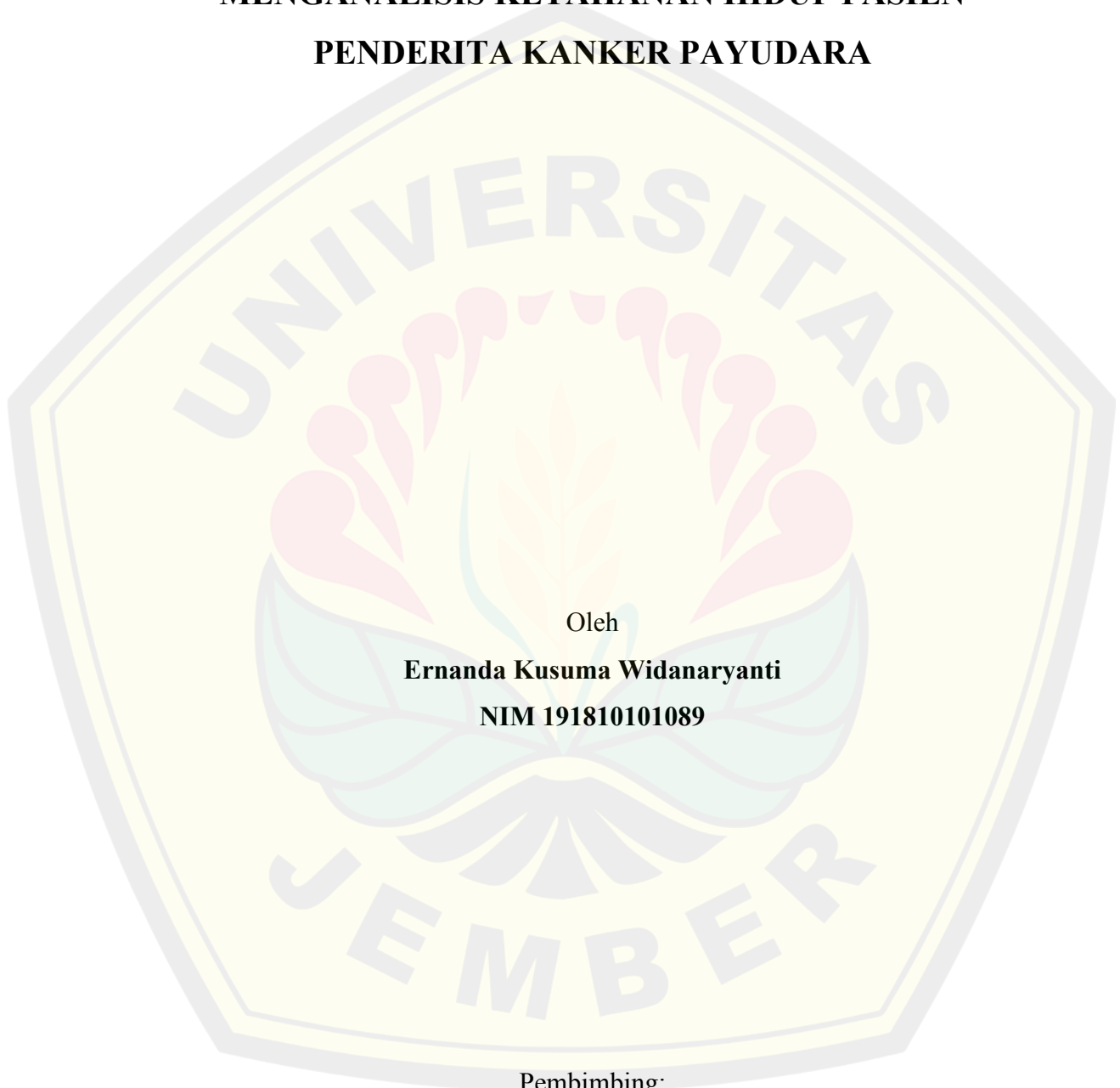


Ernanda Kusuma Widanaryanti

NIM 191810101089

SKRIPSI

PERBANDINGAN METODE *COX PROPORTIONAL HAZARD* DAN *RANDOM SURVIVAL FOREST* UNTUK MENGANALISIS KETAHANAN HIDUP PASIEN PENDERITA KANKER PAYUDARA



Oleh

Ernanda Kusuma Widanaryanti

NIM 191810101089

Pembimbing:

Dosen Pembimbing Utama : Dr. Mohamat Fatekurohman, S.Si., M.Si.

Dosen Pembimbing Anggota : Prof. Drs. I Made Tirta, M.Sc., Ph.D.

PENGESAHAN

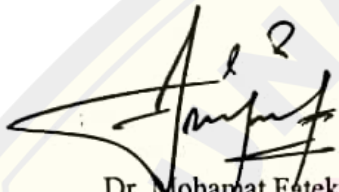
Skripsi berjudul “Perbandingan Metode *Cox Proportional Hazard* dan *Random Survival Forest* Untuk Menganalisis Ketahanan Hidup Pasien Penderita Kanker Payudara” telah diuji dan disahkan pada:

Hari, Tanggal : RABU 17 MAY 2023

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

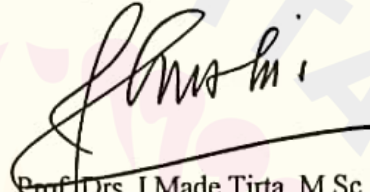
Tim Penguji:

Ketua,



Dr. Mohamat Fatekurohman, S.Si., M.Si.
NIP. 196906061998031001

Anggota I,



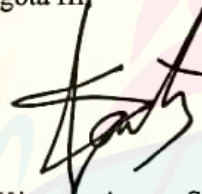
Prof. Drs. I Made Tirta, M.Sc., Ph.D.
NIP. 195912201985031002

Anggota II,



Dian Anggraeni, S.Si., M.Si.
NIP. 198202162006042002

Anggota III,



Dr. Kiswara Agung Santoso, S.Si., M.Kom.
NIP. 197209071998031003

Mengesahkan,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Jember



Drs. Achmad Saifullah, M.Sc., Ph.D.
NIP. 195910091986021001

RINGKASAN

Perbandingan Metode *Cox Proportional Hazard* dan *Random Survival Forest* Untuk Menganalisis Ketahanan Hidup Pasien Penderita Kanker Payudara; Ernanda Kusuma Widanaryanti, 191810101089; 2022: 52 halaman; Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyakit kanker menjadi masalah kesehatan di Indonesia maupun dunia. *Global Burden of Cancer Study* (GLOBOCAN) dari *World Health Organization* (WHO) (2020) mencatat bahwa total kasus kanker di Indonesia pada tahun 2020 mencapai 396.914 kasus dan total kematian sebesar 234.511 kasus. Kanker payudara memiliki jumlah kasus baru tertinggi di Indonesia sebesar 65.858 kasus atau 16,6% dari total kasus kanker. Kanker payudara (*carcinoma mammae*) adalah kanker yang terbentuk dari sel-sel ganas yang tumbuh dan berkembang tanpa dikehakui aktivitasnya sehingga dapat menyebar ke jaringan dan organ di sekitar payudara. Kanker payudara memiliki jumlah penderita tertinggi kedua di Indonesia pada wanita setelah kanker serviks.

Menurut Kleinbaum dan Klein (2012), analisis *survival* adalah suatu metode yang digunakan untuk menganalisis data dengan variabel hasil berupa waktu sampai terjadinya suatu kejadian (*event*). Analisis *survival* terbagi menjadi tiga model yaitu, model parametrik, model semiparametrik dan model nonparametrik. Metode yang paling umum digunakan dalam menganalisis kelangsungan hidup suatu individu salah satunya pada saat mengeksplorasi prognosis dalam berbagai jenis kanker yaitu metode *Cox Proportional Hazard* (CPH). Metode CPH efisien digunakan apabila asumsi *proportional hazard* terpenuhi, jika tidak terpenuhi maka model CPH tidak memberikan kesimpulan akurat sehingga diperlukan metode alternatif. Metode alternatif yang dapat digunakan salah satunya dengan metode *Machine Learning* (ML) yakni *Random Survival Forest* (RSF).

Qomaria dkk. (2019), membahas mengenai Regresi *Cox Proportional Hazard* untuk menganalisis faktor-faktor yang memengaruhi waktu *survival* atau waktu ketahanan hidup pasien *stroke*. Ghodratollah dkk. (2020), membahas mengenai analisis *survival* menggunakan metode *Random Survival Forest* untuk

mengidentifikasi faktor yang memengaruhi kelangsungan hidup pasien kanker kolorektal di Iran berdasarkan variabel demografis dan klinis terkait. Xianxin dkk. (2020), membahas mengenai perbandingan metode *Random Survival Forest* (RSF) dan metode *Cox Proportional Hazard* (CPH) yang diimplementasikan pada kasus tumor glioma derajat tinggi setelah radioterapi. Chmielewska dkk. (2020), membandingkan RSF dan CPH dalam memprediksi kebangkrutan Usaha Kecil dan Menengah (UKM) di Polandia menggunakan akurasi prediksi *C-index*.

Penelitian ini bertujuan untuk mengidentifikasi variabel yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup penderita kanker payudara di Rumah Sakit dr. Soebandi Jember. Faktor-faktor yang diduga penyebab kanker payudara adalah usia, indeks massa tubuh, riwayat penyakit lain, hemoglobin, leukosit, trombosit, penggunaan obat, disiplin kemoterapi, penggunaan dosis *cyclospamid* dan penggunaan dosis *epirubicin*. Hasil perbandingan model CPH dan RSF dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara menunjukkan bahwa metode RSF lebih baik daripada CPH dengan nilai *C-Index* masing-masing metode RSF dan CPH yaitu 0,15 dan 0,97. Faktor-faktor yang memengaruhi ketahanan hidup pasien penderita kanker payudara pada model CPH yaitu usia, indeks massa tubuh, trombosit dan penggunaan obat. Sedangkan pada model RSF, faktor-faktor yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara yaitu penggunaan obat, disiplin kemoterapi, leukosit, riwayat penyakit lain, penggunaan dosis *cyclospamid*, penggunaan dosis *epirubicin* dan hemoglobin.

PRAKATA

Puji Syukur kepada Tuhan Yang Maha Esa atas karunia dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Perbandingan Metode *Cox Proportional Hazard* dan *Random Survival Forest* Untuk Menganalisis Ketahanan Hidup Pasien Penderita Kanker Payudara” yang merupakan salah satu syarat untuk menyelesaikan pendidikan S1 pada Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember.

Dalam penyusunan skripsi ini, penulis mendapatkan banyak bimbingan, dukungan serta bantuan dari berbagai pihak, oleh sebab itu penulis dengan ketulusan dan kerendahan hati mengucapkan terimakasih kepada:

1. Dr. Mohamat Fatekurohman, S.Si., M.Si., selaku Dosen Pembimbing Utama dan Prof. Drs. I Made Tirta, M.Sc., Ph.D., selaku Dosen Pembimbing Anggota yang dengan penuh kesabaran membimbing, mengarahkan, memberikan saran dan petunjuk dalam penyusunan skripsi;
2. Dian Anggraeni, S.Si., M.Si selaku Dosen Penguji I dan Dr. Kiswara Agung Santoso, S.Si., M.Kom selaku Dosen Penguji II yang telah memberikan kritik dan saran yang membangun dalam penyusunan skripsi;
3. Dr. Mohamat Fatekurohman, S.Si., M.Si., selaku Dosen Pembimbing Akademik yang selalu membantu dan memberikan arahan dalam proses belajar selama menjadi mahasiswa pada masa kuliah;
4. Seluruh staf pengajar Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember yang telah memberikan ilmu serta bimbingannya sehingga penulis dapat menyelesaikan skripsi ini;
5. Ayahanda Kusnadi, Ibunda Ida Khotimah, Adik tercinta Erinda Kusuma Larasati serta keluarga besar saya yang selalu memberikan semangat dan memberikan do'a;
6. Ibu Wiwid, Ibu Dinda, Mas Muji beserta seluruh staf RSD dr, Soebandi yang telah membantu dalam mendapatkan data penelitian sehingga penulis dapat menyelesaikan skripsi ini;

7. Guru-guru yang telah memberikan ilmu pengetahuan dan do'a selama ini;
8. Sahabat Weny Andika Damayanti selaku teman pertama selama menjadi mahasiswa yang telah memberikan semangat dan senantiasa mendengarkan keluh kesah;
9. Teman - teman Calculus '19 Jurusan Matematika Fakultas MIPA yang telah memberi dukungan;
10. Teman - teman Kost Muslimah Halmahera, GenBI Jember, SEN Eight dan Verechelen atas semangat yang diberikan;
11. Semua pihak yang telah membantu dalam penulisan skripsi yang tidak bisa disebutkan satu persatu.

Kritik dan saran yang membangun penulis terima dari pembaca, penulis berharap skripsi ini dapat bermanfaat.

Jember, April 2023

Penulis

DAFTAR ISI

	Halaman
HALAMAN SAMPUL.....	i
HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	ii
HALAMAN MOTTO	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PEMBIMBINGAN.....	v
HALAMAN PENGESAHAN.....	vi
RINGKASAN	vii
PRAKATA.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN	xvi
BAB 1. PENDAHULUAN	2
1.1 Latar Belakang.....	2
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
BAB 2. TINJAUAN PUSTAKA.....	5
2.1 Kanker Payudara.....	5
2.1.1 Pengertian Kanker Payudara.....	5
2.1.2 Faktor Risiko Kanker Payudara	5
2.2 Analisis <i>Survival</i>	8
2.2.1 Pengertian Analisis <i>Survival</i>	8
2.2.2 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i>	9
2.3 Penyensoran Data	10
2.4 <i>Cox Proportional Hazard</i>.....	11
2.5 Pengujian Signifikansi Parameter.....	13

2.5.1 Pengujian Secara Serentak	13
2.5.2 Pengujian Secara Parsial	13
2.6 <i>Random Survival Forest</i>	13
2.7 Pemilihan Variabel <i>Importance</i>	15
2.8 <i>Harrell's Concordance Index (C-Index)</i>.....	16
BAB 3. METODOLOGI PENELITIAN	18
3.1 Data	18
3.2 Langkah-Langkah Penelitian	20
BAB 4. HASIL DAN PEMBAHASAN	23
4.1 Hasil.....	23
4.1.1 Analisis Deskriptif	23
4.1.2 Nilai Fungsi <i>Survival</i> dan <i>Hazard</i> Setiap Variabel.....	29
4.1.3 Uji Asumsi <i>Proportional Hazard</i>	41
4.1.4 Estimasi Parameter Model <i>Cox Proportional Hazard</i>	42
4.1.5 Analisis <i>Survival</i> Menggunakan Metode <i>Random Survival Forest</i>	44
4.2 Pembahasan.....	47
BAB 5. PENUTUP.....	50
5.1 Kesimpulan.....	50
5.2 Saran	50
DAFTAR PUSTAKA	51
LAMPIRAN.....	53

DAFTAR TABEL

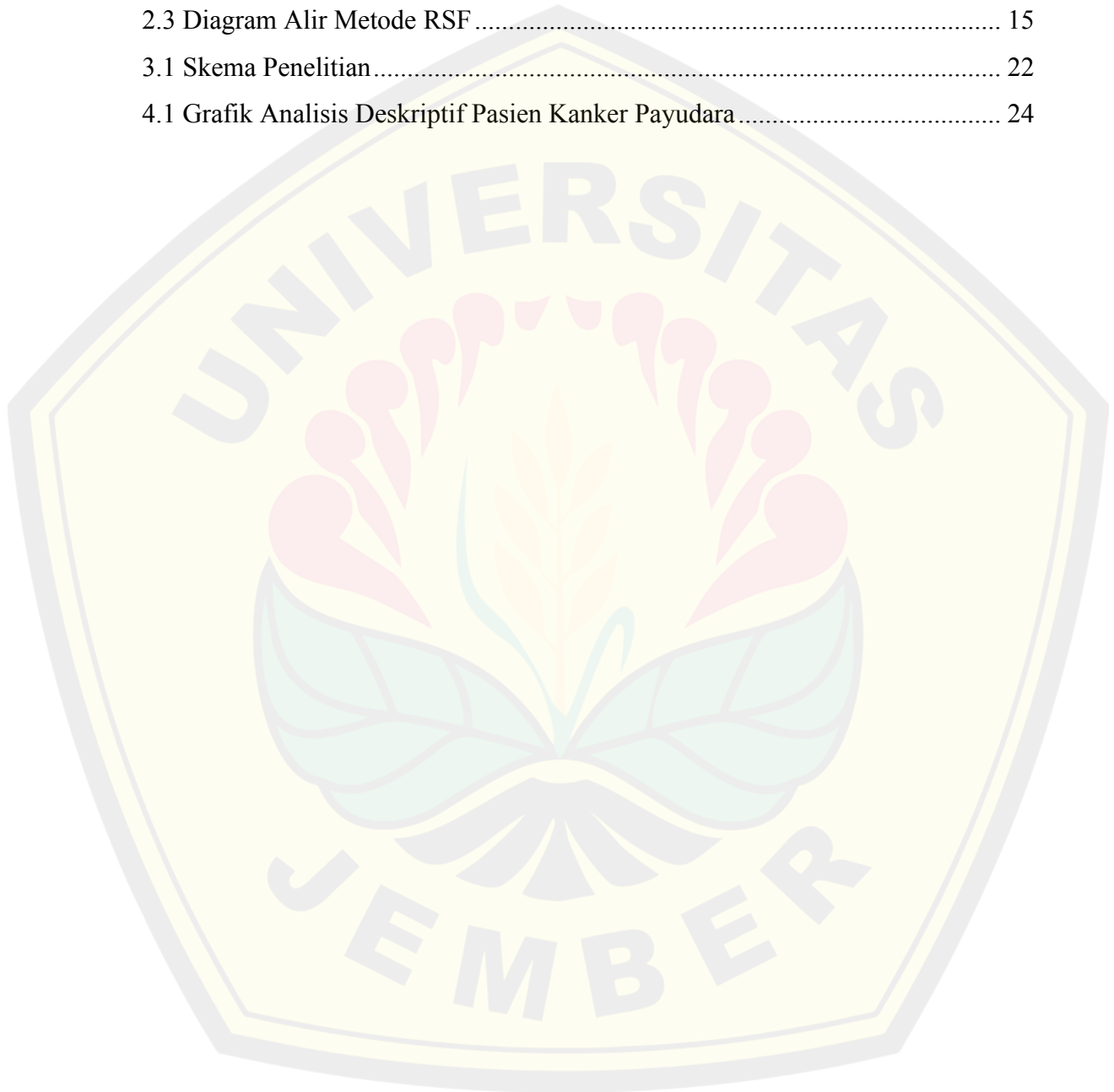
	Halaman
4.1 Analisis deskriptif pasien kanker payudara.....	23
4.2 Analisis Deskriptif Variabel Usia	24
4.3 Analisis Deskriptif Variabel IMT	25
4.4 Analisis Deskriptif Variabel RPL	25
4.5 Analisis Deskriptif Variabel HB	26
4.6 Analisis Deskriptif Variabel Leukosit.....	26
4.7 Analisis Deskriptif Variabel Trombosit.....	27
4.8 Analisis Deskriptif Variabel PO	27
4.9 Analisis Deskriptif Variabel Kemoterapi.....	28
4.10 Analisis Deskriptif Variabel Penggunaan Dosis <i>Epirubicin</i>	28
4.11 Analisis Deskriptif Variabel Penggunaan Dosis <i>Cycloposпамid</i>	29
4.12 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien $40 \leq \text{usia} < 50$	30
4.13 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien $50 \leq \text{usia} < 60$	30
4.14 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien $60 \leq \text{usia} < 70$	31
4.15 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien $18,5 \leq \text{IMT} < 25,1$	31
4.16 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien $25,1 \leq \text{IMT} < 30$	32
4.17 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Tidak Ada RPL	33
4.18 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Ada RPL	33
4.19 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> pasien HB rendah	34
4.20 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> pasien Leukosit Rendah	35
4.21 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> pasien Leukosit Tinggi.....	35
4.22 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> pasien Trombosit Rendah.....	36
4.23 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> pasien Trombosit Tinggi	36
4.24 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Obat Tetap dan Dosis Tetap	37
4.25 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Obat Diganti.....	38
4.26 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Tidak Mengikuti Kemoterapi... 39	39
4.27 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Tidak Menggunakan Dosis <i>Epirubicin</i>	40

4.28 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i> Pasien Tidak Menggunakan Dosis <i>Cycloposamid</i>	41
4.29 Nilai <i>Schoenfeld Residuals</i>	42
4.30 Hasil Estimasi Parameter Model <i>Cox Proportional Hazard</i>	42
4.31 Hasil <i>Output</i> Pengolahan Data dengan Metode RSF	44
4.32 Nilai <i>Importance</i> Variabel Bebas.....	46
4.33 <i>Harrell's Concordance Index</i> Untuk Perbandingan Metode	49



DAFTAR GAMBAR

	Halaman
2.1 Grafik Fungsi <i>Survival</i>	10
2.2 Contoh Waktu <i>Survival</i>	11
2.3 Diagram Alir Metode RSF	15
3.1 Skema Penelitian	22
4.1 Grafik Analisis Deskriptif Pasien Kanker Payudara	24



DAFTAR LAMPIRAN

	Halaman
1 Data Pasien Kanker Payudara.....	52
2 Pemanggilan data.....	53
3 <i>Script dan Output</i> Program R Analisis Deskriptif.....	53
4 <i>Script dan Output</i> Fungsi <i>Survival</i>	58
5 <i>Script dan Output</i> Uji Asumsi <i>Cox Proportional Hazard</i>	63
6 <i>Script dan Output</i> Uji Estimasi Parameter <i>Cox Proportional Hazard</i>	63
7 <i>Script dan Output</i> Program R <i>Random Survival Forest</i>	65
8 <i>Script dan Output</i> <i>C-Index</i>	66



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Penyakit kanker menjadi masalah kesehatan di Indonesia maupun dunia. Menurut Kementerian Kesehatan Republik Indonesia (2016b), kanker atau *neoplasia maligna* adalah suatu penyakit tidak menular yang ditandai dengan pertumbuhan dan perkembangan sel/jaringan yang tidak normal yang dapat menyebar sehingga merusak bagian tubuh lainnya serta dapat menyebabkan kematian. Ancaman penyakit kanker yang semakin meningkat disebabkan karena adanya perubahan pola hidup pada masyarakat. Diketahui jenis-jenis kanker yang banyak menyebabkan kematian adalah kanker payudara, paru-paru, lambung, hati, esofagus dan kolorektal.

Global Burden of Cancer Study (GLOBOCAN) dari *World Health Organization* (WHO) (2020) mencatat bahwa total kasus kanker di Indonesia pada tahun 2020 mencapai 396.914 kasus dan total kematian sebesar 234.511 kasus. Kanker payudara memiliki jumlah kasus baru tertinggi di Indonesia sebesar 65.858 kasus atau 16,6% dari total kasus kanker. Kanker payudara pada saat ini menjadi jenis penyakit yang sangat menakutkan bagi setiap perempuan di seluruh dunia, termasuk Indonesia. Berdasarkan Kementerian Kesehatan Republik Indonesia (2016a), estimasi jumlah penderita kanker payudara terbanyak terdapat pada Provinsi Jawa Timur dan Jawa Tengah. Umumnya, penyakit kanker payudara terdiagnosa atau ditemui setelah mencapai stadium lanjut. Ketahanan hidup pasien penderita kanker payudara dipengaruhi oleh berbagai faktor, sehingga biasanya penderita penyakit ini hanya bertahan hidup dalam waktu yang relatif singkat. Secara statistik, waktu ketahanan hidup pasien kanker payudara dapat dianalisis dengan menggunakan analisis *survival*.

Menurut Kleinbaum dan Klein (2012), analisis *survival* adalah suatu metode yang digunakan untuk menganalisis data dengan variabel hasil berupa waktu sampai terjadinya suatu kejadian (*event*). Waktu ketahanan hidup atau waktu *survival* dapat didefinisikan sebagai waktu dari awal observasi hingga terjadinya

peristiwa yang dapat berupa hari, bulan, maupun tahun. Analisis *survival* terbagi menjadi tiga model yaitu, model parametrik, model semiparametrik dan model nonparametrik. Metode yang paling umum digunakan dalam menganalisis kelangsungan hidup suatu individu salah satunya pada saat mengeksplorasi prognosis dalam berbagai jenis kanker yaitu metode *Cox Proportional Hazard* (CPH). *Cox Proportional Hazard* adalah metode statistik yang dikenal baik untuk mengeksplorasi korelasi antara waktu *survival* dan kovariat. Metode CPH efisien digunakan apabila asumsi *proportional hazard* terpenuhi, jika tidak terpenuhi maka model CPH tidak memberikan kesimpulan akurat sehingga diperlukan metode alternatif. Metode alternatif yang dapat digunakan salah satunya dengan metode *Machine Learning* (ML) yakni *Random Survival Forest* (RSF). Menurut Xianxin dkk. (2020), metode RSF merupakan metode *ensemble* yang digunakan untuk menganalisis kelangsungan hidup tersensor kanan tanpa memerhatikan asumsi apapun.

Penelitian sebelumnya yang dilakukan oleh Qomaria dkk. (2019), membahas mengenai Regresi *Cox Proportional Hazard* untuk menganalisis faktor-faktor yang memengaruhi waktu *survival* atau waktu ketahanan hidup pasien *stroke*. Faktor-faktor yang memengaruhi ketahanan hidup pasien *stroke* adalah usia, status *diabestes mellitus*, dan jenis *stroke*. Ghodratollah dkk. (2020), membahas mengenai analisis *survival* menggunakan metode *Random Survival Forest* untuk mengidentifikasi faktor yang memengaruhi kelangsungan hidup pasien kanker kolorektal di Iran berdasarkan variabel demografis dan klinis terkait. Kesimpulan dari penelitian tersebut adalah empat prediktor yang paling penting dari kelangsungan hidup untuk pasien kanker kolorektal yaitu metastasis ke organ lain, jumlah WBC, stadium penyakit dan jumlah limfoma. Xianxin dkk. (2020), membahas mengenai perbandingan metode *Random Survival Forest* (RSF) dan metode *Cox Proportional Hazard* (CPH) yang diimplementasikan pada kasus tumor glioma derajat tinggi setelah radioterapi. Hasil dari penelitian ini yaitu model CPH menunjukkan kinerja yang lebih baik. Chmielewska dan Matuszyk (2020), membandingkan RSF dan CPH dalam memprediksi kebangkrutan Usaha Kecil dan Menengah (UKM) di Polandia menggunakan akurasi prediksi *C-index*. Hasil

penelitian menunjukkan bahwa RSF tidak hanya memberikan hasil yang lebih baik tetapi juga lebih stabil daripada model CPH semiparametrik. Pada beberapa penelitian terdahulu, tidak ada penelitian yang mengeksplorasi antara CPH dan RSF untuk menganalisis ketahanan hidup pasien penderita kanker payudara sehingga penulis tergerak untuk melakukan penelitian tersebut.

Berdasarkan uraian tersebut, metode analisis *survival* yang digunakan dalam penelitian ini adalah *Cox Proportional Hazard* dan *Random Survival Forest*. Kinerja dari kedua metode ini akan dibandingkan untuk mengetahui manakah metode yang lebih baik untuk mengidentifikasi variabel yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara dengan menggunakan pendekatan *C-index* untuk mendapatkan hasil yang akurat. Data yang digunakan merupakan data pasien penderita kanker payudara yang dirawat di RSD dr. Soebandi Jember.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana hasil perbandingan dan pemilihan metode terbaik antara metode *Cox Proportional Hazard* dan *Random Survival Forest* dalam mengidentifikasi variabel yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara di RSD dr. Soebandi Jember ?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengetahui hasil perbandingan antara metode *Cox Proportional Hazard* dan *Random Survival Forest* dalam mengidentifikasi variabel yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara di RSD dr. Soebandi Jember.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini sebagai berikut:

1. Dapat memberikan informasi dan peningkatan pengetahuan kepada pembaca tentang peran statistik dalam bidang kesehatan.
2. Dapat menginformasikan kepada tenaga medis dalam menangani pasien kanker payudara agar lebih memperhatikan faktor-faktor yang memengaruhi ketahanan hidup pasien penderita kanker payudara.



BAB 2. TINJAUAN PUSTAKA

Pada bab 2 ini akan disajikan mengenai tinjauan pustaka yang berkaitan dengan kanker payudara, analisis *survival*, penyensoran data, *Cox Proportional Hazard*, uji signifikansi parameter, *Random Survival Forest*, pemilihan variabel *importance* dan *Herrell's Concordance Index (C-Index)*.

2.1 Kanker Payudara

2.1.1 Pengertian Kanker Payudara

Menurut Kementerian Kesehatan Republik Indonesia (2016b), kanker payudara (*carcinoma mammae*) adalah kanker yang terbentuk dari sel-sel ganas yang tumbuh dan berkembang tanpa dikehakui aktivitasnya sehingga dapat menyebar ke jaringan dan organ di sekitar payudara. Kanker payudara memiliki jumlah penderita tertinggi kedua di Indonesia pada wanita setelah kanker serviks. Penyebab terjadinya kanker payudara karena terganggunya sistem proliferasi sel pada jaringan payudara. Sel abnormal yang tumbuh perlahan di sekitar area tersebut bisa merusak sistem payudara (Nurcahyo, 2010). Pertumbuhan sel kanker payudara dipengaruhi oleh beberapa hormon, salah satunya yaitu hormon estrogen. Hormon estrogen tersebut dapat menyebabkan pertumbuhan sel abnormal pada organ tertentu. Selain itu, tingginya paparan hormon estrogen disebabkan oleh banyak faktor antara lain seperti tidak menyusui, tidak pernah melahirkan atau melahirkan pertama kali di atas usia 35 tahun, menstruasi di bawah usia 12 tahun, lama menstruasi diatas usia 50 tahun dan pemakaian kontrasepsi hormonal dalam jangka panjang (Suryaningsih dan Sukaca, 2009).

2.1.2 Faktor Risiko Kanker Payudara

Menurut (Kaminska, 2015), faktor risiko yang berpengaruh terhadap ketahanan hidup pasien kanker payudara yaitu:

1. Usia

Faktor risiko yang diketahui sangat penting pada kasus kanker payudara yaitu usia. Kanker payudara telah didiagnosis sekitar 80% terjadi pada wanita pasca menopause (Rebolj dkk., 2015). Kanker payudara dapat meningkat cepat pada usia reproduktif. Wanita dengan usia lebih dari 30 tahun memiliki kemungkinan besar dapat menderita kanker payudara dan akan terus meningkat hingga usia 50 tahun ataupun setelah *menopause*. Sekitar 18% penderita kanker payudara terdiagnosa pada wanita dengan usia 40 tahun dan 77% terdiagnosa pada wanita dengan usia 50 tahun. Usia pada penderita kanker payudara lebih muda dibandingkan dengan penderita kanker paru-paru.

2. Jenis Kelamin

Secara umum, wanita memiliki sel/jaringan pada payudara lebih banyak dibandingkan laki-laki sehingga wanita memiliki risiko lebih besar daripada laki-laki untuk terkena kanker payudara. Kasus kanker payudara kemungkinan banyak dipengaruhi oleh hormon-hormon pada wanita seperti hormon estrogen dan progesteron yang berpengaruh terhadap proses proliferasi sel-sel pada kelenjar payudara yang secara fisiologis juga lebih berkembang dibandingkan laki-laki. Laki-laki juga dapat terkena kanker payudara tetapi penyakit ini 100 kali lebih sering ditemukan pada wanita.

3. Faktor Reproduksi

Wanita yang tidak pernah melahirkan atau pernah melahirkan pertama kali pada usia lebih dari 30 tahun memiliki risiko lebih besar terkena kanker payudara dibandingkan dengan wanita yang melahirkan pada usia kurang dari 30 tahun. Kehamilan pertama pada usia kurang dari 18 tahun memiliki risiko setengah dari wanita yang hamil pada usia lebih dari 30 tahun. Wanita yang hamil pertama kali pada usia lebih dari 40 tahun berisiko 3 kali lebih besar dibandingkan wanita yang hamil pada usia kurang dari 40 tahun.

4. Riwayat Keluarga/Faktor Genetik

Diketahui apabila terdapat anggota dalam satu keluarga yang menderita penyakit kanker maka dapat dinyatakan bahwa itu merupakan penyakit kanker yang familial. Sekitar 75% dari sindroma tersebut disebabkan adanya mutasi pada gen

p53 yang merupakan gen penekan tumor (*suppressor gene*). Seseorang dapat berisiko terkena kanker payudara lebih besar apabila anggota keluarganya ada yang menderita kanker payudara atau kanker ovarium. Diperkirakan 15%-20% kanker payudara dihubungkan dengan adanya riwayat pada keluarga (faktor genetik). Wanita yang memiliki riwayat kanker payudara pada keluarga tingkat pertama (ibu/saudara perempuan) memiliki risiko 2-3 kali lebih besar terkena kanker payudara.

5. Riwayat Penyakit Pada Payudara

Wanita yang pernah atau sedang menderita kelainan proliferasi memiliki risiko lebih besar terkena kanker payudara. Adanya kanker payudara kontralateral dan wanita dengan riwayat pernah mengalami infeksi dan operasi tumor jinak pada payudara memiliki risiko 3-9 kali lebih besar.

6. Status Obesitas

Indikator utama pada status obesitas adalah Indeks Massa Tubuh (IMT) dengan rasio berat badan (kilogram) dan kuadrat tinggi badan (meter). Menurut hasil penelitian Dietz dkk. (2012), mengenai hubungan antara ukuran tubuh dengan risiko terkena kanker payudara tergantung pada usia antara wanita premenopausal dan postmenopausal. Usia sebelum menopause, risiko terkena kanker payudara lebih rendah dibandingkan pada wanita postmenopause karena peningkatan berat badan 10 kg akan berpengaruh \pm 80% meningkatkan risiko mengalami kanker payudara. Wanita yang mengalami obesitas cenderung lebih berisiko terkena kanker payudara. Olahraga dan pola makan merupakan upaya yang baik yang dapat mengontrol kadar hormon plasma yang memengaruhi perkembangan kanker payudara pada wanita yang mengalami obesitas.

7. Kebiasaan Merokok

Wanita dengan kebiasaan merokok akan memiliki tingkat metabolisme estrogen lebih tinggi dibandingkan wanita yang tidak merokok. Pada wanita dengan riwayat keluarga penderita kanker payudara atau kanker ovarium, kebiasaan merokok ini dapat meningkatkan risiko terkena kanker payudara sebesar 2,4 kali dibandingkan dengan yang tidak merokok.

8. Jenis Pengobatan

Ketahanan hidup penderita kanker payudara dapat ditinjau dari pengobatan awalnya. Ketahanan hidup penderita kanker payudara yang tidak melakukan pengobatan sebesar 35%. Ada empat cara yang dapat digunakan untuk pengobatan kanker payudara yaitu, terapi bedah, radioterapi, kemoterapi dan bioterapi. Keempat terapi harus digunakan secara kombinasi untuk mendapatkan hasil yang maksimal. Kemoterapi adalah penanganan paling utama yang dapat dilakukan pada penderita kanker payudara (Wulandari, 2012). Kemoterapi saat ini menjadi cara yang paling sering digunakan oleh pasien kanker payudara setelah ataupun sebelum melakukan terapi lainnya.

9. Status Sosial Ekonomi

Status sosial ekonomi menengah ke bawah secara tidak langsung diduga memiliki peningkatan risiko terkena kanker payudara. Tingkat sosial ekonomi seperti pendidikan, pendapatan dan pengetahuan memengaruhi seseorang dalam gaya hidup (*life style*) dan perilaku hidup sehat seseorang. Hal ini berkaitan dengan kebiasaan merokok dan minuman beralkohol, kurangnya aktivitas fisik serta kurangnya konsumsi buah dan sayuran.

10. Paparan Radiasi

Wanita yang memiliki riwayat terkena paparan radiasi dosis tinggi terutama selama masa remaja, berisiko lebih besar terkena kanker payudara. Risiko kanker payudara akibat radiasi dipengaruhi oleh dosis radiasi yang diterima, umur pada saat terpapar radiasi, lama paparan, jenis kelamin dan faktor genetik.

2.2 Analisis Survival

2.2.1 Pengertian Analisis *Survival*

Menurut Kleinbaum dan Klein (2012), analisis *survival* adalah suatu metode statistik yang digunakan untuk menganalisis data sehingga didapatkan variabel hasil, dimana variabel hasil tersebut adalah waktu (*time*) sampai terjadinya suatu peristiwa (*event*) tertentu atau disebut dengan waktu *survival*. Waktu *survival* merupakan suatu variabel dependen dalam analisis *survival* yang mengukur waktu awal penelitian (*time origin*) yang telah ditentukan sampai dengan terjadinya suatu

peristiwa tertentu (*event*) atau waktu akhir pengamatan (*end point*). Berbagai peristiwa khusus yang dapat terjadi dalam kehidupan sehari-hari seperti kematian, aktivitas penyakit yang muncul setelah pengobatan atau timbulnya penyakit terbaru, respon terhadap eksperimen atau aktivitas berbeda yang diputuskan sesuai dengan tujuan peneliti. Terdapat 3 faktor dalam penentuan waktu *survival* seperti berikut ini:

1. Waktu awal pencatatan (*start-point*) adalah waktu dimulainya pencatatan untuk setiap individu pada suatu peristiwa.
2. Waktu akhir pencatatan (*end-point*) adalah akhir dari pencatatan untuk menentukan status tersensor atau tidak tersensor untuk dilakukan analisis.
3. Skala pengukuran digambarkan sebagai batasan waktu kejadian dari awal hingga akhir kejadian, misalnya skala harian, mingguan, dan tahunan.

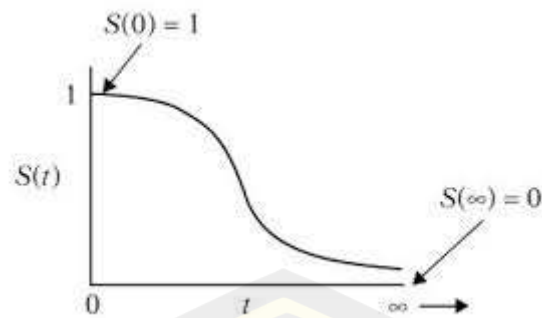
2.2.2 Fungsi Survival dan Fungsi Hazard

Analisis *survival* terbagi menjadi dua fungsi yang sering digunakan yaitu fungsi *survival* dan fungsi *hazard* (Kleinbaum & Klein, 2012). Fungsi ketahanan hidup atau fungsi *survival* $S(t)$ yang merupakan komplemen dari fungsi distribusi kumulatif $F(t)$. Fungsi ketahanan didefinisikan dengan peluang suatu individu untuk dapat bertahan hidup berdasarkan waktu *survival* sampai dengan waktu t . Hal ini juga diartikan sebagai peluang suatu variabel acak T lebih dari waktu t seperti persamaan 2.1 berikut ini. Fungsi *survival* $S(t)$ didefinisikan:

$$S(t) = 1 - F(t) = P(T > t) \quad (2.1)$$

Fungsi ketahanan atau fungsi *survival* memiliki tiga sifat, yaitu:

- a. Fungsi ketahanan atau fungsi *survival* menyempit searah dengan bertambahnya taraf t ,
- b. Nilai $P(T > 0) = 1$: pada awal penelitian subjek belum mengalami kegagalan, dengan nilai $t = 0, S(t) = S(0) = 1$,
- c. Nilai $t = \infty, S(t = \infty) = 0$: jangka waktu penelitian diperpanjang tanpa batas, tidak terdapat subjek yang bertahan.

Gambar 2.1 Grafik Fungsi *Survival*

(Sumber: Harlan, 2017)

Fungsi *hazard* adalah kelajuan suatu individu mengalami kejadian dalam interval waktu t sampai $t + \Delta t$ dengan ketentuan individu masih bertahan hidup hingga waktu ke- t . Fungsi *hazard* ditanyakan dalam persamaan (2.3)

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2.2)$$

dengan T merupakan variabel acak non-negatif ($T \geq 0$) yang menyatakan waktu *survival* objek dari suatu populasi.

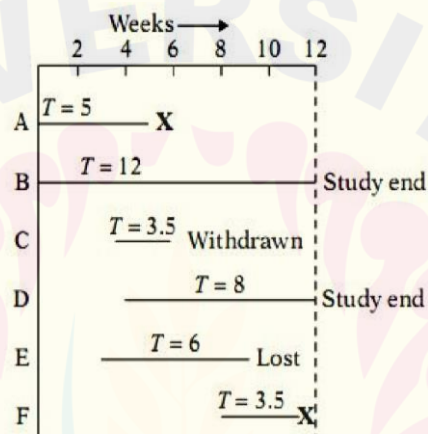
2.3 Penyensoran Data

Penyensoran data merupakan proses untuk mengatasi ketidaklengkapan data pengamatan. Jika waktu *survival* hanya diamati sebagian dan tidak sampai pada *failure time* maka dapat dikatakan data tersensor. Menurut Kleinbaum dan Klein (2012), terdapat beberapa penyebab terjadinya data tersensor sebagai berikut:

- Loss to follow up*, ketika objek yang diamati tidak mengalami suatu *event* sampai berakhirnya penelitian, seperti meninggal, menolak untuk berpartisipasi atau pindah.
- Drop out*, terjadi ketika perlakuan dihentikan karena beberapa alasan tertentu.
- Termination of study*, terjadi saat objek yang diamati dan diobservasi tapi terjadi *failure time*.

Menurut Menurut Kleinbaum dan Klein (2012), terdapat tiga jenis penyensoran pada analisis *survival* sebagai berikut:

1. Sensor kanan (*right censoring*), terjadi apabila suatu objek yang masuk dalam penelitian bisa diamati namun tidak mengalami kejadian hingga akhir penelitian atau tidak bisa mengikuti penelitian sampai akhir akibat adanya kejadian diluar yang menjadi penelitian.
2. Sensor kiri (*left censoring*), terjadi apabila pada waktu awal suatu objek belum diamati namun hingga sebelum penelitian berakhir kejadian sudah dapat teramati secara penuh.
3. Sensor interval (*interval censoring*) merupakan sensor yang waktu *survival*nya berada di dalam selang waktu tertentu.



Gambar 2.2 Contoh Waktu *Survival*
(Kleinbaum & Klein, 2012)

Pada Gambar 2.1 menunjukkan pencatatan kejadian dari awal pencatatan hingga akhir waktu pencatatan. Skala waktu ditentukan dalam mingguan dan *failure* yang berbeda-beda pada pencatatan masing-masing individu. Contohnya pada kejadian A dan F mengalami *event* sampai sebelum berakhirnya penelitian. Pada kejadian B dan D, tidak mengalami *event* sampai berakhirnya penelitian. Sedangkan pada kejadian C dan E, perlakuan dihentikan atau hilang sebelum penelitian berakhir karena beberapa alasan tertentu.

2.4 *Cox Proportional Hazard*

Cox Proportional Hazard diperkenalkan oleh Cox (1972) merupakan suatu pendekatan model semiparametrik yang paling umum digunakan dalam

menevaluasi variabel independen dengan waktu ketahanan, dimana pada model ini tidak memerlukan adanya informasi mengenai distribusi yang mendasari waktu ketahanan (Kleinbaum dan Klein, 2012). Model *Cox Proportional Hazard* dikatakan sebagai model *Cox* karena asumsi *proportional hazard* atau rasio dari fungsi tingkat kegagalan suatu individu satu dengan individu lainnya adalah konstan. Model *Cox Proportional Hazard* dapat dituliskan dengan persamaan

$$\begin{aligned} h(t|\mathbf{X}) &= h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p) \\ &= h_0(t) \exp\left(\sum_{i=1}^p \beta_i X_i\right) \end{aligned} \quad (2.3)$$

dengan

$h(t|\mathbf{X})$: fungsi *hazard* pada waktu t bagi suatu individu tertentu dengan faktor risiko \mathbf{X}

$h_0(t)$: fungsi *baseline hazard* pada waktu t

β_i : parameter regresi dari variabel – variabel bebas yang terdiri dari $\beta_1, \beta_2, \dots, \beta_n$

X_i : variabel – variabel bebas yang terdiri dari X_1, X_2, \dots, X_n

t : waktu *survival*

(Harlan, 2017).

Baseline hazard merupakan fungsi *hazard* yang bergantung pada waktu dan tidak melibatkan variabel-variabel prediktor atau kovariatnya. Model Regresi *Cox Proportional Hazard* ini bisa digunakan dalam mengetahui hubungan antara waktu kelangsungan hidup individu dengan variabel yang diduga memengaruhinya. Model *Cox Proportional Hazard* yang diterapkan pada data *survival*, data yang digunakan untuk model ini harus memenuhi asumsi *proportional hazard* (Harlan, 2017).

2.5 Pengujian Signifikansi Parameter

Pengujian parameter pada analisis data *survival* menurut Hosmer *et al.* (2011) terbagi menjadi dua yaitu :

2.5.1 Pengujian Secara Serentak

Pengujian secara serentak yang akan digunakan sebagai berikut.

Hipotesis :

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

$$H_1 : \text{minimal ada satu dari } \beta_k \neq 0, \text{ dengan } k = 1, 2, \dots, p$$

Taraf Signifikansi : $\alpha = 0,05$

Statistik ujinya adalah sebagai berikut

$$G^2 = -2 \ln \frac{L(\hat{\omega})}{L(\hat{\Omega})} \quad (2.4)$$

Keterangan :

$L(\hat{\omega})$: nilai *likelihood* untuk model tanpa menyertakan variabel predictor.

$L(\hat{\Omega})$: nilai *likelihood* untuk model dengan menyertakan variabel prediktor.

Tolak H_0 , jika $G^2 \leq \chi_{\alpha,p}^2$ atau $p - value < \alpha$

2.5.2 Pengujian Secara Parsial

Pengujian secara parsial yang akan digunakan adalah

Hipotesis :

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0; k = 1, 2, 3, \dots, p$$

Taraf Signifikansi : $\alpha = 0.05$

Statistik ujinya adalah sebagai berikut

$$W^2 = \frac{\hat{\beta}_k^2}{(SE(\hat{\beta}_k))^2} \quad (2.5)$$

Tolak H_0 jika $W^2 > \chi_{\alpha,p}^2$ atau $p - value < \alpha$

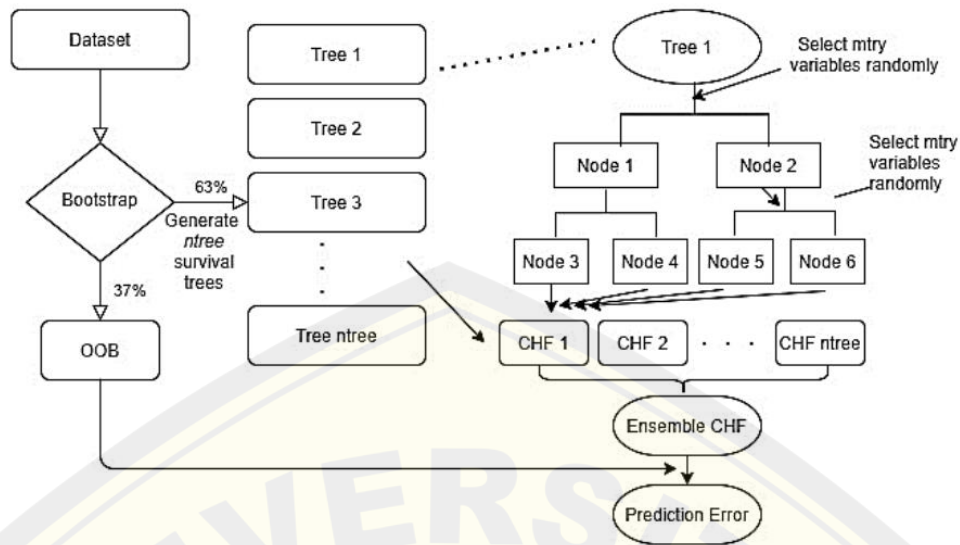
2.6 *Random Survival Forest*

Random Survival Forest (RSF) diperkenalkan oleh Ishwaran *et al.* (2008) merupakan metode pendekatan *machine learning* yang paling umum digunakan dan sebagai salah satu metode pohon *ensemble* dalam menganalisis data *survival time-*

to-event tersensor kanan (Xianxin dkk., 2020). RSF merupakan perluasan dari metode *Random Forest* (RF) yang diperkenalkan oleh Breiman (2001). Implementasi algoritma *Random Survival Forest* terdiri dari langkah-langkah berikut ini:

1. Ambil sebanyak B contoh *bootstrap* dari data dengan melakukan pengembalian. Setiap contoh *bootstrap* digunakan untuk membentuk pohon *survival*. Setiap sampel *bootstrap* umumnya mencakup sekitar 63% dari data asli dan dilakukan penyisihan terhadap sekitar 37% dari data, yang disebut dengan *Out-of-Bag* (OOB).
2. Tumbuhkan pohon kelangsungan hidup (*survival tree*) untuk setiap sampel *bootstrap*. Pada setiap simpul pohon, pilih secara acak p variabel kandidat. Node dipecah menggunakan variabel kandidat yang memaksimalkan perbedaan kelangsungan hidup antara node anak. yang melibatkan waktu *survival* dan informasi status sensor digunakan.
3. Tumbuhkan pohon ke ukuran penuh dengan batasan bahwa yang harus dimiliki oleh *node terminal* minimal $d_0 > 0$ data kegagalan unik.
4. Hitung nilai fungsi kumulatif hazard (CHF) untuk setiap pohon. Kemudian cari nilai CHF ansambel dengan cara mencari nilai rata-rata dari seluruh pohon dalam *forests*.
5. Hitung kesalahan prediksi untuk CHF ansambel menggunakan data OOB. Kemudian kesalahan prediksi yang didapat akan digunakan untuk memilih variabel penting.

Algoritma *Random Survival Forest* diuraikan dalam *flowchart* sebagai berikut



Gambar 2.3 Diagram Alir Metode RSF

(Sumber: Ishwaran, 2008)

2.7 Pemilihan Variabel *Importance*

Variabel dapat dipilih dengan memfilter kepentingan variabelnya atau *Variable Importance* (VIMP). VIMP merupakan perbedaan dalam kesalahan prediksi OOB sebelum dan sesudah permutasi, sehingga nilai VIMP yang besar menunjukkan bahwa kesalahan spesifikasi mengurangi akurasi prediksi pada hutan. VIMP dengan nilai kepentingan yang positif menunjukkan bahwa variabel tersebut memiliki kemampuan prediksi. Sementara itu jika VIMP mendekati nol menunjukkan peubah tidak memberikan kontribusi apa pun terhadap akurasi prediksi, dan nilai negatif menunjukkan akurasi prediksi meningkat ketika terdapat kesalahan dalam menentukan peubah. Perhitungan VIMP dilakukan pada seluruh hutan acak (Saadati dan Bagheri 2019). VIMP dihitung dengan membandingkan data OOB dengan pohon *survival* yang dibangun sebelumnya. Oleh sebab itu VIMP mengukur perubahan kesalahan prediksi ketika peubah yang diinginkan tidak tersedia saat membangun hutan baru.

2.8 Harrell's Concordance Index (C-Index)

Menurut Ishwaran dkk. (2008), evaluasi metrik yang paling populer yang digunakan dalam menentukan menilai akurasi kinerja prediksi model pada analisis *survival* tersensor kanan yaitu Harrell's Concordance index (C-index). C-Index mengukur apakah suatu peristiwa yang terjadi lebih awal dapat dikaitkan dengan tingkat risiko yang lebih tinggi atau tingkat ketahanan yang lebih rendah. Langkah-langkah menghitung kesalahan prediksi dengan C-Index sebagai berikut:

1. Setiap pasangan pasien i dan j (dengan $i \neq j$), lihat skor risiko (η) dan waktu kejadian (T).
2. Jika T_i dan T_j tidak tersensor, maka dapat mengamati kapan kedua pasien terkena penyakit. Pasangan (i, j) adalah *concordant pair* jika $\eta_i > \eta_j$ dan $T_i < T_j$, dan itu adalah *discordant pair* jika $\eta_i > \eta_j$ dan $T_i > T_j$.
3. Jika T_i dan T_j tersensor, maka tidak diketahui siapa yang terkena penyakit pertama. Jadi, tidak mempertimbangkan pasangan ini dalam perhitungan.
4. Jika salah satu T_i dan T_j tersensor, maka hanya dapat mengamati satu penyakit. Katakanlah mengamati pasien yang mendapatkan penyakit pada waktu T_i , dan bahwa T_j tersensor. (Logika yang sama berlaku untuk situasi sebaliknya).
 - Jika $T_j < T_i$, maka tidak diketahui pasti siapa yang terkena penyakit ini terlebih dahulu, jadi tidak mempertimbangkan pasangan ini dalam perhitungan.
 - Jika $T_j > T_i$, maka dapat diketahui pasti siapa pasien yang terkena penyakit ini pertama. Oleh karena itu, (i, j) adalah *concordant pair* jika $\eta_i > \eta_j$ dan merupakan *discordant pair* jika $\eta_i < \eta_j$.
5. Nilai C-Index (C) dinyatakan sebagai berikut

$$C = \frac{\text{concordant pairs}}{\text{concordant pairs} + \text{discordant pairs}} \quad (2.6)$$

C-index ekuivalen dengan tingkat kesalahan prediksi (*error rate*), yang mana tingkat kesalahan prediksi dinyatakan sebagai berikut

$$\text{Error rate} = 1 - C, \quad 0 \leq \text{Error} \leq 1 \quad (2.7)$$

$C = 1$ atau *Error rate* = 0 menunjukkan prediksi sempurna, sedangkan $C =$ *Error rate* = 0,5 menunjukkan sesuai prosedur yang tidak lebih baik dari tebakan acak (Ishwaran, 2008).



BAB 3. METODOLOGI PENELITIAN

Pada bab 3 ini akan disajikan mengenai modologi penelitian yang berkaitan pemilihan jenis data, variabel penelitian, langkah-langkah dan *flowchart* penelitian.

3.1 Data

Jenis data yang akan digunakan pada penelitian ini yaitu data sekunder. Data sekunder tersebut merupakan data pasien kanker payudara yang dirawat di RSD dr. Soebandi Jember. Periode pengamatan dalam penelitian ini dilakukan dari bulan Januari-Januari 2023, adapun variabel-variabelnya meliputi

a) Variabel dependen (Y) pada penelitian ini yaitu lama waktu ketahanan hidup pasien penderita kanker payudara di Rumah Sakit (T) dan status pasien (D) dengan penjelasan sebagai berikut

1) Lama Waktu Ketahanan Hidup Pasien (T)

Penetapan lama waktu ketahanan hidup pasien kanker payudara terhitung dari pertama kali dinyatakan menderita hingga terjadinya *event* yaitu meninggal. Variabel lama waktu ketahanan hidup pasien kanker payudara merupakan data kontinu dengan skala rasio.

2) Status Pasien (D)

Penetapan status pasien penderita kanker payudara berskala nominal dengan dua kategori yaitu

0 : pasien kanker payudara tidak meninggal, menolak untuk berpartisipasi, pindah Rumah Sakit lain atau meninggal dengan karena alasan tertentu (tersensor)

1 : pasien kanker payudara yang dinyatakan meninggal (tidak tersensor)

3) Variabel independen (X) merupakan variabel yang akan digunakan pada penelitian ini antara lain

1) Usia (X_1)

0 : $30 \leq X_1 < 40$

- 1 : $40 \leq X_1 < 50$
2 : $50 \leq X_1 < 60$
3 : $60 \leq X_1 < 70$
4 : $X_1 \geq 70$
- 2) Indeks Massa Tubuh (X_2)
0 : Kurus ($X_2 < 18,5$)
1 : Normal ($18,5 \leq X_2 < 25,1$)
2 : Gemuk ($25,1 \leq X_2 < 30$)
3 : Obesitas ($X_2 \geq 30$)
- 3) Riwayat Penyakit Lain (X_3)
0 : Tidak ada
1 : Ada
- 4) HB (X_4)
0 : HB rendah ($X_4 < 12$)
1 : HB normal ($12 \leq X_4 < 16,1$)
2 : HB tinggi ($X_4 > 16$)
- 5) Leukosit (X_5)
0 : Leukosit rendah ($4000 > X_5$)
1 : Leukosit normal ($4000 \leq X_5 < 10001$)
2 : Leukosit tinggi ($X_5 > 10000$)
- 6) Trombosit (X_6)
0 : Trombosit rendah ($X_6 < 2.10^5$)
1 : Trombosit normal ($2.10^5 \leq X_6 \leq 4.10^5$)
2 : Trombosit tinggi ($X_6 > 4.10^5$)
- 7) Penggunaan Obat (X_7)
0 : Obat tetap dan dosis tetap
1 : Obat tetap dengan dosis berbeda
2 : Obat diganti
- 8) Disiplin Kemoterapi (X_8)
0 : Mengikuti kemoterapi

1 : Tidak mengikuti kemoterapi

9) Penggunaan Dosis *Epirubicin* (X_9)

0 : Tidak menggunakan

1 : Dosis tetap

2 : Dosis berubah

10) Penggunaan Dosis *Cyclospamid* (X_{10})

0 : Tidak menggunakan

1 : Dosis tetap

2 : Dosis berubah

3.2 Langkah-Langkah Penelitian

Penelitian tentang “Perbandingan Metode *Cox Proportional Hazard* dan *Random Survival Forest* untuk Menganalisis Ketahanan Hidup Pasien Penderita Kanker Payudara di Kabupaten Jember“ dengan langkah-langkah sebagai berikut.

1. Studi literatur

Studi literatur yang dilakukan berupa pendekatan secara teoritis dengan mengkaji teori yang berkaitan dengan analisis *survival* khususnya mengenai metode *Cox Proportional Hazard* dan *Random Survival Forest*.

2. Pengumpulan data

Proses pengambilan data dilakukan secara sekunder. Data sekunder tersebut merupakan data pasien kanker payudara yang dirawat di rumah sakit di Kabupaten Jember

3. Analisis terhadap variabel

Melakukan analisis deskriptif terhadap variabel terikat yaitu waktu, dan variabel bebas berupa faktor-faktor yang diduga dapat memprediksi ketahanan hidup pasien penderita kanker payudara.

4. Membandingkan hasil kinerja metode

Proses membandingkan hasil kinerja dua metode, yaitu metode *Cox Proportional Hazard* dan *Random Survival Forest* dilakukan dengan menggunakan pendekatan *C-Index*. Metode yang memiliki nilai *Error C* –

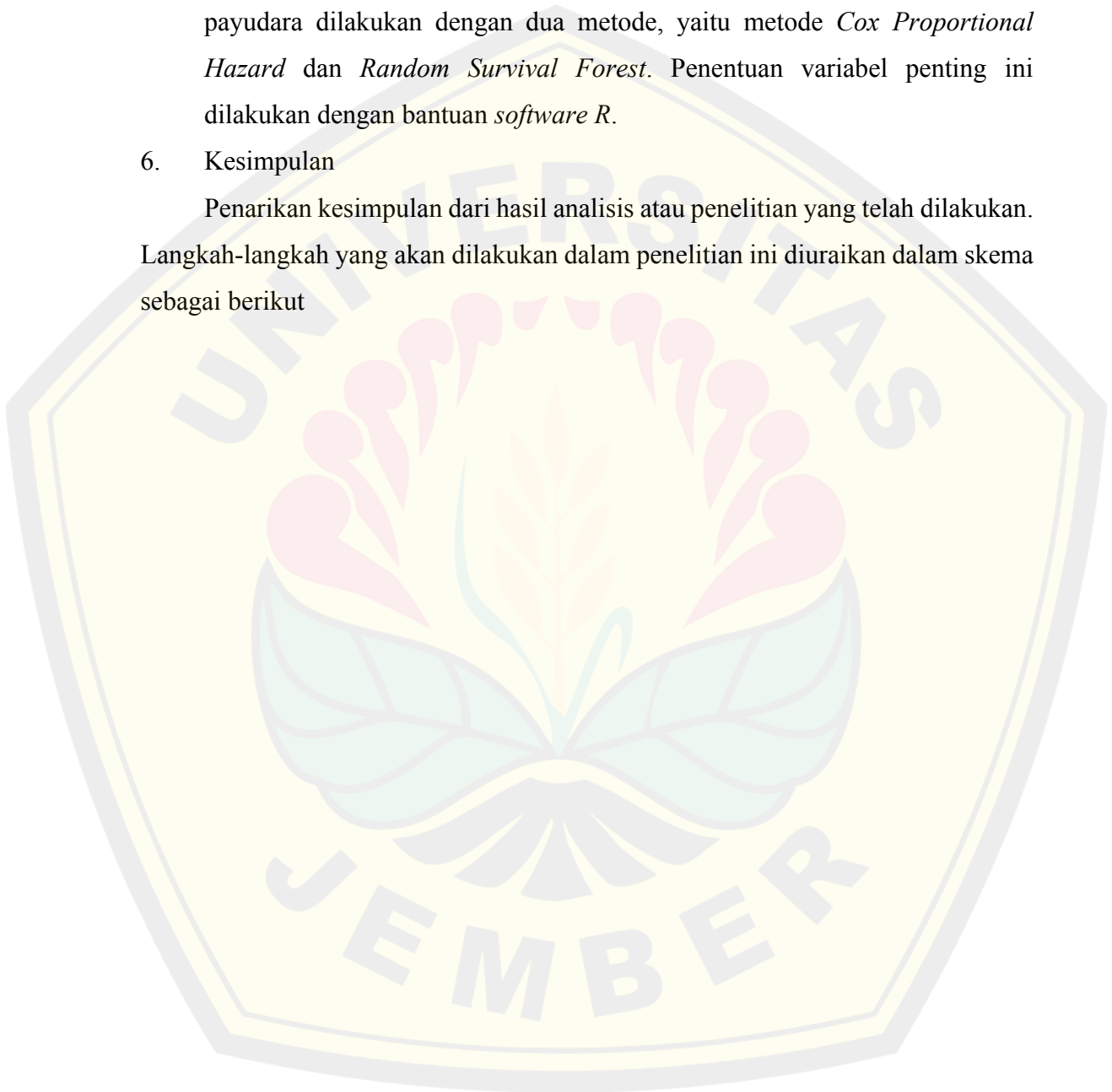
Index yang lebih rendah dapat dikatakan bahwa metode tersebut lebih baik atau lebih cocok digunakan pada data pasien penderita kanker payudara.

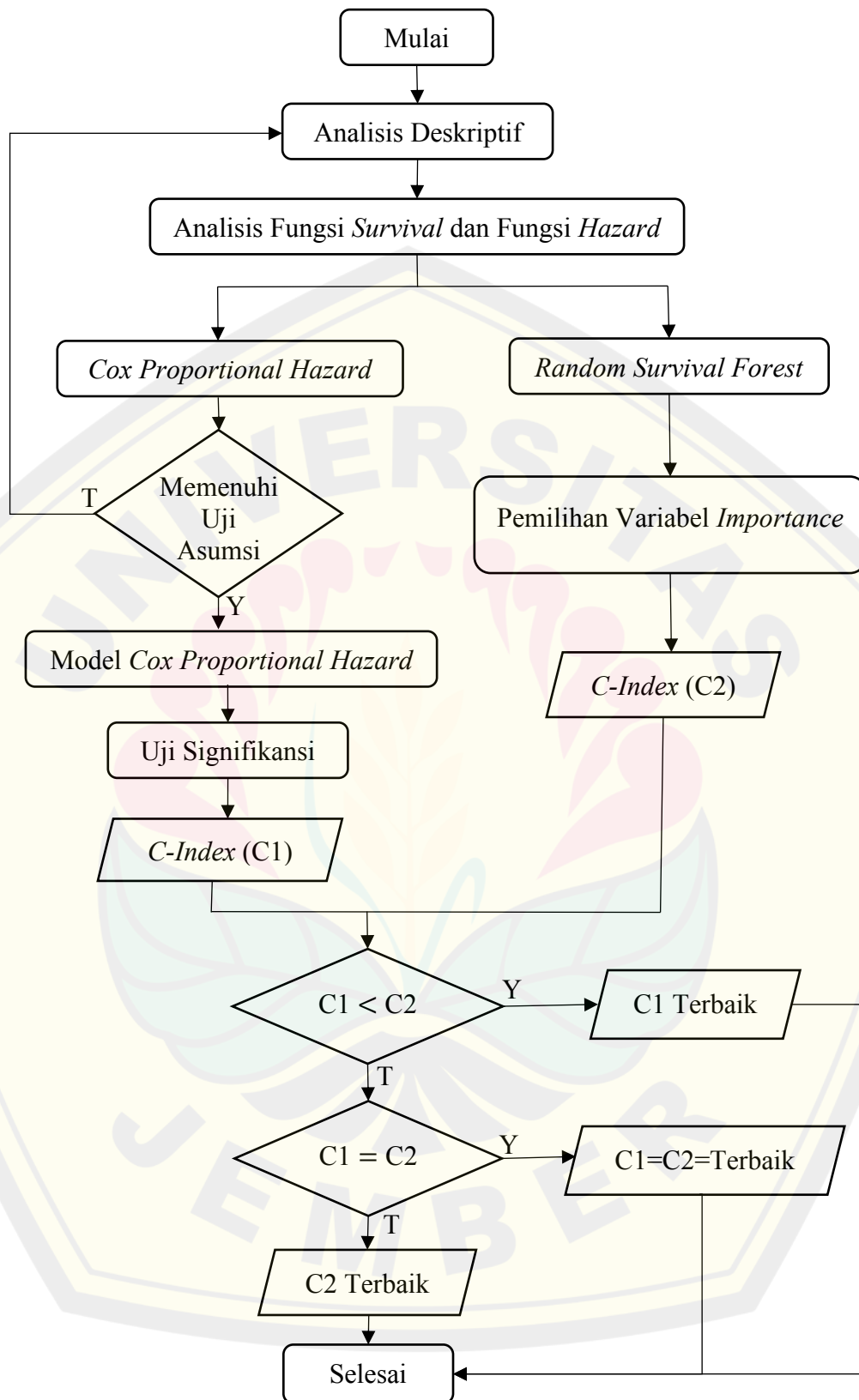
5. Menentukan variabel *importance*

Menentukan variabel-variabel penting atau variabel-variabel yang secara signifikan dapat memprediksi ketahanan hidup pasien penderita kanker payudara dilakukan dengan dua metode, yaitu metode *Cox Proportional Hazard* dan *Random Survival Forest*. Penentuan variabel penting ini dilakukan dengan bantuan *software R*.

6. Kesimpulan

Penarikan kesimpulan dari hasil analisis atau penelitian yang telah dilakukan. Langkah-langkah yang akan dilakukan dalam penelitian ini diuraikan dalam skema sebagai berikut





Gambar 3.1 Skema Penelitian

BAB 4. HASIL DAN PEMBAHASAN**4.1 Hasil**

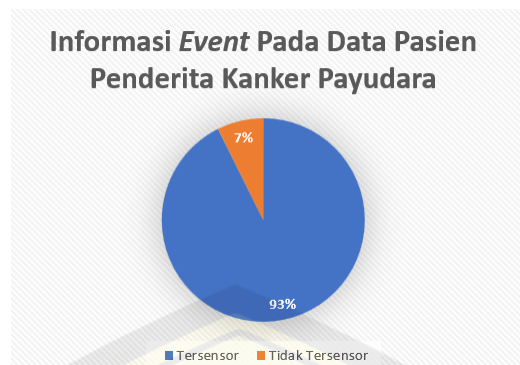
4.1.1 Analisis Deskriptif

Pada penelitian ini akan dilakukan analisis statistika deskriptif terlebih dahulu dari pasien penderita kanker payudara yang dijadikan sampel penelitian sebelum melakukan pembahasan inti. Analisis statistika deskriptif bertujuan untuk mengetahui gambaran secara umum mengenai sifat atau karakteristik data sampel. Data yang digunakan pada penelitian ini sebanyak 68 data pasien penderita kanker payudara di RSD dr. Soebandi Jember berupa data sekunder. *Event* kegagalan analisis *survival* pada ketahanan hidup pasien penderita kanker payudara adalah kejadian ketika pasien dinyatakan meninggal. Data *event* yang dihasilkan terbagi menjadi data tersensor dan tidak tersensor (teramati) seperti pada Tabel 4.1.

Tabel 4.1 Analisis Deskriptif Pasien Kanker Payudara

Informasi <i>Event</i>	Frekuensi	Persentase
Tersensor	63	93%
Tidak tersensor (teramati)	5	7%
Jumlah	68	100%

Tabel 4.1 menunjukkan bahwa persentase *event* tersensor atau pasien kanker payudara yang masih hidup sebanyak 93% atau 63 pasien. Persentase *event* teramati atau pasien kanker payudara yang meninggal sebanyak 7% atau 5 pasien. Pasien tersensor dinyatakan dengan “0” dan pasien yang tidak tersensor dinyatakan dengan “1”. Adapun grafik dari analisis deskriptif data pasien penderita kanker payudara dapat dilihat pada Gambar 4.1.



Gambar 4.1 Grafik Analisis Deskriptif Pasien Kanker Payudara

Data memiliki 10 variabel yang merupakan faktor-faktor yang diduga menjadi penyebab pasien mengalami kanker payudara. 10 variabel yang digunakan dalam penelitian ini yaitu usia (X_1), indeks massa tubuh (X_2), riwayat penyakit lain (X_3), hemoglobin (X_4), leukosit (X_5), trombosit (X_6), penggunaan obat (X_7), disiplin kemoterapi (X_8), penggunaan dosis *epirubicin* (X_9) dan penggunaan dosis *cyclospamid* (X_{10}). Hasil analisis deskriptif yang diperoleh dari program R untuk setiap variabel yang diteliti adalah sebagai berikut:

a. Variabel Usia

Analisis deskriptif pasien penderita kanker payudara yang menjalani rawat inap di RSD dr. Soebandi Jember berdasarkan variabel usia dapat dilihat pada Tabel 4.2.

Tabel 4.2 Analisis Deskriptif Variabel Usia

Kategori Usia	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
$30 \leq \text{usia} < 40$	7	7	0	0	0
$40 \leq \text{usia} < 50$	18	15	3	0,1667	0
$50 \leq \text{usia} < 60$	23	22	1	0,04348	0
$60 \leq \text{usia} < 70$	17	16	1	0,05882	0
$\text{usia} \geq 70$	3	3	0	0	0

Jumlah pasien penderita kanker payudara dengan kategori $40 \leq \text{usia} < 50$ mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,1667.

b. Variabel Indeks Massa Tubuh (IMT)

Hasil *output* analisis deskriptif dari program R pada variabel indeks massa tubuh dapat dilihat pada Tabel 4.3.

Tabel 4.3 Analisis Deskriptif Variabel IMT

Kategori IMT	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
$IMT < 18,5$	5	5	0	0	0
$18,5 \leq IMT < 25,1$	38	35	3	0,07895	0
$25,1 \leq IMT < 30$	15	13	2	0,1333	0
$IMT \geq 30$	10	10	0	0	0

Jumlah pasien penderita kanker payudara dengan kategori $25,1 \leq IMT < 30$ mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,1333.

c. Variabel Riwayat Penyakit Lain (RPL)

Hasil *output* analisis deskriptif dari program R pada variabel riwayat penyakit lain dapat dilihat pada Tabel 4.4.

Tabel 4.4 Analisis Deskriptif Variabel RPL

Kategori RPL	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
Tidak ada	41	40	1	0,02439	0
Ada	27	23	4	0,1481	0

Jumlah pasien penderita kanker payudara dengan kategori ada riwayat penyakit lain mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,1481.

d. Variabel HB

Hasil *output* analisis deskriptif dari program R pada variabel HB dapat dilihat pada Tabel 4.5.

Tabel 4.5 Analisis Deskriptif Variabel HB

Kategori HB	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
HB rendah	42	37	5	0,119	0
HB normal	25	25	0	0	0
HB tinggi	1	1	0	0	0

Jumlah pasien penderita kanker payudara dengan kategori HB rendah mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,119.

e. Variabel Leukosit

Hasil *output* analisis deskriptif dari program R pada variabel leukosit dapat dilihat pada Tabel 4.6.

Tabel 4.6 Analisis Deskriptif Variabel Leukosit

Kategori Leukosit	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
Leukosit rendah	6	4	2	0,3333	0
Leukosit normal	33	33	0	0	0
Leukosit tinggi	29	26	3	0,1034	0

Jumlah pasien penderita kanker payudara dengan kategori leukosit rendah mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,3333.

f. Variabel Trombosit

Hasil *output* analisis deskriptif dari program R pada variabel trombosit dapat dilihat pada Tabel 4.7.

Tabel 4.7 Analisis Deskriptif Variabel Trombosit

Kategori	Jumlah	Status		Mean	Median
		Hidup	Meninggal		
Trombosit rendah	Pasien 12	9	3	(hari) 0,25	(hari) 0
Trombosit normal	37	37	0	0	0
Trombosit tinggi	19	17	2	0,1053	0

Jumlah pasien penderita kanker payudara dengan kategori trombosit rendah mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,25.

g. Variabel Penggunaan Obat (PO)

Hasil *output* analisis deskriptif dari program R pada variabel penggunaan obat dapat dilihat pada Tabel 4.8.

Tabel 4.8 Analisis Deskriptif Variabel PO

Kategori	Jumlah	Status		Mean	Median
		Hidup	Meninggal		
PO Obat tetap dan dosis tetap	Pasien 56	54	2	(hari) 0,03571	(hari) 0
Obat tetap dengan dosis berubah	6	6	0	0	0
	6	3	3	0,5	0,5

Obat
diganti

Jumlah pasien penderita kanker payudara dengan kategori obat diganti mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,5.

h. Variabel Disiplin Kemoterapi (DK)

Hasil *output* analisis deskriptif dari program R pada variabel disiplin kemoterapi dapat dilihat pada Tabel 4.9.

Tabel 4.9 Analisis Deskriptif Variabel Kemoterapi

Kategori Kemoterapi	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
Mengikuti Kemoterapi	33	33	0	0	0
Tidak Mengikuti Kemoterapi	35	30	5	0,1429	0

Jumlah pasien penderita kanker payudara dengan kategori tidak mengikuti kemoterapi mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,1429.

i. Variabel Penggunaan Dosis *Epirubicin*

Hasil *output* analisis deskriptif dari program R pada variabel penggunaan dosis *epirubicin* dapat dilihat pada Tabel 4.10.

Tabel 4.10 Analisis Deskriptif Variabel Penggunaan Dosis *Epirubicin*

Kategori Dosis <i>Epirubicin</i>	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
Tidak menggunakan	58	53	5	0,08621	0

Dosis Tetap	10	10	0	0	0
Dosis Berubah	0	0	0	0	0

Jumlah pasien penderita kanker payudara dengan kategori tidak menggunakan dosis *epirubicin* mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,08621.

j. Variabel Penggunaan Dosis *Cyclospamid*

Hasil *output* analisis deskriptif dari program R pada variabel penggunaan dosis *cyclospamid* dapat dilihat pada Tabel 4.11.

Tabel 4.11 Analisis Deskriptif Variabel Penggunaan Dosis *Cyclospamid*

Kategori Dosis <i>Cyclospamid</i>	Jumlah Pasien	Status		Mean (hari)	Median (hari)
		Hidup	Meninggal		
Tidak menggunakan Dosis Tetap	40	35	5	0,125	0
Dosis Berubah	25	25	0	0	0
	3	3	0	0	0

Jumlah pasien penderita kanker payudara dengan kategori tidak menggunakan dosis *cyclospamid* mempunyai proporsi meninggal yang lebih tinggi dibandingkan dengan kategori lainnya yaitu 0,125.

4.1.2 Nilai Fungsi *Survival* dan *Hazard* Setiap Variabel

Probabilitas variabel selama waktu t dapat diketahui melalui perhitungan nilai *survival* dan *hazard* setiap variabelnya dengan waktu *survival* dalam hari. Semua variabel dikategorikan untuk mempermudah dalam mengambil kesimpulan keproporsionalan *hazard* dari hasil nilai *survival* dan *hazard* masing-masing variabel.

a. Variabel Usia

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu $30 \leq \text{usia} < 40$, $40 \leq \text{usia} < 50$, $50 \leq \text{usia} < 60$, $60 \leq \text{usia} < 70$, dan $\text{usia} \geq 70$, dapat diketahui

dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.12, 4.13 dan 4.14.

Tabel 4.12 Fungsi *Survival* dan Fungsi *Hazard* Pasien $40 \leq \text{usia} < 50$

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	18	1	0,944	0,944	0,056	0,058
28	16	1	0,938	0,885	0,115	0,122
55	10	1	0,9	0,797	0,203	0,227

Pada Tabel 4.12 untuk pasien dengan kategori $40 \leq \text{usia} < 50$ diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,944 artinya pasien dengan kategori $40 \leq \text{usia} < 50$ mempunyai ketahanan hidup sebesar 94,4% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori $40 \leq \text{usia} < 50$ yaitu 18 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,944. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,058, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 5,8%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,056 artinya risiko kematian pertama adalah 5,6%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

Tabel 4.13 Fungsi *Survival* dan Fungsi *Hazard* Pasien $50 \leq \text{usia} < 60$

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
67	12	1	0,916	0,916	0,084	0,088

Pada Tabel 4.13 untuk pasien dengan kategori $50 \leq \text{usia} < 60$ diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,916 artinya pasien dengan kategori $50 \leq \text{usia} < 60$ mempunyai ketahanan hidup sebesar 91,6% pada waktu *survival* pertama (67 hari). Jumlah pasien dengan kategori $50 \leq \text{usia} < 60$ yaitu 12 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,916. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,088, artinya risiko kematian kumulatif dari

waktu *survival* pertama adalah 8,8%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,084 artinya risiko kematian pertama adalah 8,4%.

Tabel 4.14 Fungsi *Survival* dan Fungsi *Hazard* Pasien $60 \leq \text{usia} < 70$

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
157	1	1	0	0	1	NA

Pada Tabel 4.14 untuk pasien dengan kategori $60 \leq \text{usia} < 70$ diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0 artinya pasien dengan kategori $60 \leq \text{usia} < 70$ mempunyai ketahanan hidup sebesar 0% pada waktu *survival* pertama (157 hari). Jumlah pasien dengan kategori $60 \leq \text{usia} < 70$ yaitu 1 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh NA, artinya risiko kematian kumulatif dari waktu *survival* pertama tidak dapat diprediksi. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 1 artinya risiko kematian pertama adalah 100%.

b. Variabel Indeks Massa Tubuh (IMT)

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu $\text{IMT} < 18,5$, $18,5 \leq \text{IMT} < 25,1$, $25,1 \leq \text{IMT} < 30$, dan $\text{IMT} \geq 30$, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.15 dan 4.16.

Tabel 4.15 Fungsi *Survival* dan Fungsi *Hazard* Pasien $18,5 \leq \text{IMT} < 25,1$

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
28	37	1	0,973	0,973	0,027	0,027
55	26	1	0,962	0,936	0,064	0,066
67	18	1	0,944	0,884	0,116	0,123

Pada Tabel 4.15 untuk pasien dengan kategori $18,5 \leq \text{IMT} < 25,1$ diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,973 artinya pasien dengan kategori $18,5$

$\leq \text{IMT} < 25,1$ mempunyai ketahanan hidup sebesar 97,3% pada waktu *survival* pertama (28 hari). Jumlah pasien dengan kategori $18,5 \leq \text{IMT} < 25,1$ yaitu 37 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,973. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,027, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 2,7%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,027 artinya risiko kematian pertama adalah 2,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

Tabel 4.16 Fungsi *Survival* dan Fungsi *Hazard* Pasien $25,1 \leq \text{IMT} < 30$

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	15	1	0,933	0,933	0,067	0,069
157	2	1	0,5	0,467	0,533	0,761

Pada Tabel 4.16 untuk pasien dengan kategori $25,1 \leq \text{IMT} < 30$ diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,933 artinya pasien dengan kategori $25,1 \leq \text{IMT} < 30$ mempunyai ketahanan hidup sebesar 93,3% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori $25,1 \leq \text{IMT} < 30$ yaitu 15 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,933. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,069, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 6,9%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,067 artinya risiko kematian pertama adalah 6,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

c. Variabel Riwayat Penyakit Lain (RPL)

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu tidak ada dan ada riwayat penyakit lain, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.17 dan 4.18.

Tabel 4.17 Fungsi *Survival* dan Fungsi *Hazard* Pasien Tidak Ada RPL

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
55	32	1	0,969	0,969	0,031	0,031

Pada Tabel 4.17 untuk pasien dengan kategori tidak ada riwayat penyakit lain diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,969 artinya pasien dengan kategori tidak ada riwayat penyakit lain mempunyai ketahanan hidup sebesar 96,9% pada waktu *survival* pertama (55 hari). Jumlah pasien dengan kategori tidak ada riwayat penyakit lain yaitu 32 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,969. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,031, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 3,1%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,031 artinya risiko kematian pertama adalah 3,1%.

Tabel 4.18 Fungsi *Survival* dan Fungsi *Hazard* Pasien Ada RPL

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	27	1	0,963	0,963	0,037	0,038
28	25	1	0,96	0,924	0,076	0,079
67	13	1	0,923	0,853	0,147	0,159
157	1	1	0	0	1	NA

Pada Tabel 4.18 untuk pasien dengan kategori ada riwayat penyakit lain diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,963 artinya pasien dengan kategori ada riwayat penyakit lain mempunyai ketahanan hidup sebesar 96,3% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori ada riwayat penyakit lain yaitu 27 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,963. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,038, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 3,8%. Sedangkan untuk

melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,037 artinya risiko kematian pertama adalah 3,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

d. Variabel HB

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu HB rendah, HB normal dan HB tinggi, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.19.

Tabel 4.19 Fungsi *Survival* dan Fungsi *Hazard* pasien HB rendah

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	42	1	0,976	0,976	0,024	0,024
28	39	1	0,974	0,951	0,049	0,050
55	31	1	0,968	0,920	0,08	0,083
67	24	1	0,958	0,882	0,118	0,126
157	2	1	0,5	0,441	0,559	0,819

Pada Tabel 4.19 untuk pasien dengan kategori HB rendah diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,976 artinya pasien dengan kategori HB rendah mempunyai ketahanan hidup sebesar 97,6% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori HB rendah yaitu 42 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,976. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,024, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 2,4%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,024 artinya risiko kematian pertama adalah 2,4%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

e. Variabel Leukosit

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu leukosit rendah, leukosit normal dan leukosit tinggi, dapat diketahui dari hasil *output script* program

R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.20 dan 4.21.

Tabel 4.20 Fungsi *Survival* dan Fungsi *Hazard* pasien Leukosit Rendah

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
28	6	1	0,833	0,833	0,167	0,183
55	4	1	0,75	0,625	0,375	0,470

Pada Tabel 4.20 untuk pasien dengan kategori leukosit rendah diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,833 artinya pasien dengan kategori leukosit rendah mempunyai ketahanan hidup sebesar 83,3% pada waktu *survival* pertama (28 hari). Jumlah pasien dengan kategori leukosit rendah yaitu 6 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,833. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,183, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 18,3%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,167 artinya risiko kematian pertama adalah 16,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

Tabel 4.21 Fungsi *Survival* dan Fungsi *Hazard* pasien Leukosit Tinggi

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$H(t)$
6	29	1	0,966	0,966	0,034	0,035
67	14	1	0,929	0,897	0,103	0,109
157	1	1	0	0	1	NA

Pada Tabel 4.21 untuk pasien dengan kategori leukosit tinggi diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,966 artinya pasien dengan kategori leukosit tinggi mempunyai ketahanan hidup sebesar 96,6% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori leukosit tinggi yaitu 29 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,966. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari

nilai *hazard* kumulatif, dan diperoleh 0,035, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 3,5%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,034 artinya risiko kematian pertama adalah 3,4%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

f. Variabel Trombosit

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu trombosit rendah, trombosit normal dan trombosit tinggi, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.22 dan 4.23.

Tabel 4.22 Fungsi *Survival* dan Fungsi *Hazard* pasien Trombosit Rendah

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
28	12	1	0,917	0,917	0,083	0,087
55	8	1	0,875	0,802	0,198	0,220
157	1	1	0	0	1	NA

Pada Tabel 4.22 untuk pasien dengan kategori trombosit rendah diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,917 artinya pasien dengan kategori trombosit rendah mempunyai ketahanan hidup sebesar 91,7% pada waktu *survival* pertama (28 hari). Jumlah pasien dengan kategori trombosit rendah yaitu 12 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,917. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,087, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 8,7%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,083 artinya risiko kematian pertama adalah 8,3%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

Tabel 4.23 Fungsi *Survival* dan Fungsi *Hazard* pasien Trombosit Tinggi

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	19	1	0,947	0,947	0,053	0,054

67	8	1	0,875	0,829	0,171	0,188
----	---	---	-------	-------	-------	-------

Pada Tabel 4.23 untuk pasien dengan kategori trombosit tinggi diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,947 artinya pasien dengan kategori trombosit tinggi mempunyai ketahanan hidup sebesar 94,7% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori trombosit tinggi yaitu 19 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,947. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,054, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 5,4%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,053 artinya risiko kematian pertama adalah 5,3%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

g. Variabel Penggunaan Obat (PO)

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu obat tetap dan dosis tetap, obat tetap dengan dosis berubah dan obat diganti, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.24 dan 4.25.

Tabel 4.24 Fungsi *Survival* dan Fungsi *Hazard* Pasien Obat Tetap dan Dosis

Tetap						
t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
67	33	1	0,970	0,970	0,03	0,03
157	2	1	0,5	0,485	0,515	0,724

Pada Tabel 4.24 untuk pasien dengan kategori obat tetap dan dosis tetap diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,970 artinya pasien dengan kategori obat tetap dan dosis tetap mempunyai ketahanan hidup sebesar 97% pada waktu *survival* pertama (67 hari). Jumlah pasien dengan kategori obat tetap dan dosis tetap yaitu 33 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,970. Nilai estimasi fungsi *survival* dapat

digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,03, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 3%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,03 artinya risiko kematian pertama adalah 3%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

Tabel 4.25 Fungsi *Survival* dan Fungsi *Hazard* Pasien Obat Diganti

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	6	1	0,833	0,833	0,167	0,183
28	5	1	0,8	0,667	0,333	0,405
55	3	1	0,667	0,444	0,556	0,811

Pada Tabel 4.25 untuk pasien dengan kategori obat diganti diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,833 artinya pasien dengan kategori obat diganti mempunyai ketahanan hidup sebesar 83,3% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori obat diganti yaitu 6 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,833. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,183, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 18,3%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,167 artinya risiko kematian pertama adalah 16,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

h. Variabel Disiplin Kemoterapi (DK)

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu mengikuti kemoterapi dan tidak mengikuti kemoterapi, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.26.

Tabel 4.26 Fungsi *Survival* dan Fungsi *Hazard* Pasien Tidak Mengikuti Kemoterapi

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	35	1	0,971	0,971	0,029	0,029
28	34	1	0,971	0,943	0,057	0,059
55	23	1	0,957	0,902	0,098	0,103
67	17	1	0,941	0,849	0,151	0,164
157	1	1	0	0	1	NA

Pada Tabel 4.26 untuk pasien dengan kategori tidak mengikuti kemoterapi diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,971 artinya pasien dengan kategori tidak mengikuti kemoterapi mempunyai ketahanan hidup sebesar 97,1% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori tidak mengikuti kemoterapi yaitu 35 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,971. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,029, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 2,9%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,029 artinya risiko kematian pertama adalah 2,9%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

i. Variabel Penggunaan Dosis *Epirubicin*

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu tidak menggunakan, dosis tetap dan dosis berubah pada penggunaan dosis *epirubicin*, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.27.

Tabel 4.27 Fungsi *Survival* dan Fungsi *Hazard* Pasien Tidak Menggunakan Dosis *Epirubicin*

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	58	1	0,983	0,983	0,017	0,017
28	54	1	0,981	0,965	0,035	0,036
55	38	1	0,973	0,939	0,061	0,063
67	27	1	0,963	0,904	0,096	0,1
157	1	1	0	0	1	NA

Pada Tabel 4.27 untuk pasien dengan kategori tidak menggunakan dosis *epirubicin* diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,983 artinya pasien dengan kategori tidak menggunakan dosis *epirubicin* mempunyai ketahanan hidup sebesar 98,3% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori tidak menggunakan dosis *epirubicin* yaitu 58 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,983. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,017, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 1,7%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,017 artinya risiko kematian pertama adalah 1,7%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

j. Variabel Penggunaan Dosis *Cyclospamid*

Nilai fungsi *survival* dan fungsi *hazard* dengan kategori yaitu tidak menggunakan, dosis tetap dan dosis berubah pada penggunaan dosis *cyclospamid*, dapat diketahui dari hasil *output script* program R pada lampiran 4. Adapun hasil analisis dengan kategori-kategori tersebut dapat dilihat pada Tabel 4.28.

Tabel 4.28 Fungsi *Survival* dan Fungsi *Hazard* Pasien Tidak Menggunakan Dosis *Cyclospamid*

t_j	n_j	d_j	$(n_j - d_j)/n_j$	$\hat{S}(t)$	$\hat{h}(t)$	$\hat{H}(t)$
6	40	1	0,975	0,975	0,025	0,025
28	38	1	0,974	0,949	0,051	0,052
55	27	1	0,963	0,914	0,086	0,090
67	21	1	0,952	0,871	0,129	0,138
157	2	1	0,5	0,435	0,565	0,832

Pada Tabel 4.28 untuk pasien dengan kategori tidak menggunakan dosis *cyclospamid* diperoleh nilai $\hat{S}(t)$ untuk fungsi *survival* pertama yaitu 0,975 artinya pasien dengan kategori tidak menggunakan dosis *cyclospamid* mempunyai ketahanan hidup sebesar 97,5% pada waktu *survival* pertama (6 hari). Jumlah pasien dengan kategori tidak menggunakan dosis *cyclospamid* yaitu 40 pasien yang dapat dilihat dari n_1 , adapun fungsi *survival* pertama dengan jumlah pasien meninggal 1 orang yaitu 0,975. Nilai estimasi fungsi *survival* dapat digunakan untuk mencari nilai *hazard* kumulatif, dan diperoleh 0,025, artinya risiko kematian kumulatif dari waktu *survival* pertama adalah 2,5%. Sedangkan untuk melihat nilai *hazard* setiap waktu *survival* bisa dilihat pada tabel $\hat{h}(t)$ dan diperoleh pada waktu *survival* pertama yaitu sebesar 0,025 artinya risiko kematian pertama adalah 2,5%. Untuk waktu *survival* selanjutnya mempunyai penjelasan yang sama.

4.1.3 Uji Asumsi *Proportional Hazard*

Uji asumsi *proportional hazard* dilakukan untuk menguji korelasi antar variabel dengan menggunakan *Goodness of Fit* yang digunakan adalah *Schoenfeld Residuals* dan pengujian yang digunakan adalah

H_0 : asumsi *proportional hazard* terpenuhi

H_1 : asumsi *proportional hazard* tidak terpenuhi

Pengecekan asumsi *proportional hazard* yaitu jika $p - value < 0,05$ maka hipotesis H_0 ditolak sehingga kovariat yang diuji tidak memenuhi asumsi

proportional hazard. Hasil pengujian dengan *Schoenfeld Residuals* seperti pada Tabel 4.29.

Tabel 4.29 Nilai *Schoenfeld Residuals*

Variabel	<i>Chisq</i>	Df	<i>p-value</i>	Keputusan
Usia	$1,50 \times 10^{-1}$	1	0,70	Terima H_0
IMT	$6,67 \times 10^{-1}$	1	0,41	Terima H_0
RPL	$2,10 \times 10^{-1}$	1	0,65	Terima H_0
HB	$4,38 \times 10^{-7}$	1	1,00	Terima H_0
Leukosit	$3,88 \times 10^{-1}$	1	0,53	Terima H_0
Trombosit	$4,67 \times 10^{-2}$	1	0,83	Terima H_0
PO	$1,75 \times 10^1$	1	0,19	Terima H_0
DK	$3,83 \times 10^{-6}$	1	1,00	Terima H_0
Epiru	$5,42 \times 10^{-4}$	1	0,98	Terima H_0
Cyclo	$1,16 \times 10^{-6}$	1	1,00	Terima H_0

Tabel 4.29 memperlihatkan nilai *p-value* untuk semua kovariat atau variabel bebas lebih besar dari nilai signifikansi $\alpha = 0,05$ sehingga keputusannya adalah hipotesis H_0 diterima. Hal ini berarti bahwa semua kovariat atau variabel memenuhi asumsi *proportional hazard*.

4.1.4 Estimasi Parameter Model *Cox Proportional Hazard*

Variabel bebas yang memenuhi asumsi *proportional hazard* adalah usia, IMT, RPL, HB, leukosit, trombosit, PO, DK, epiru dan cyclo. Variabel bebas yang sudah memenuhi asumsi tersebut akan dilakukan estimasi parameter model *Cox Proportional Hazard*. Hasil estimasi parameter untuk variabel bebas pada pasien penderita kanker payudara seperti pada Tabel 4.30.

Tabel 4.30 Hasil Estimasi Parameter Model *Cox Proportional Hazard*

Variabel	Coef	exp(coef)	SE(coef)	<i>p-value</i>
Usia	-2,567	0,07678	0,5464	$2,63 \times 10^{-6}$
IMT	-3,196	0,04091	1,141	0,00509

RPL	0,8062	2,239	1,280	0,52886
HB	-14,95	$3,216 \times 10^{-7}$	$2,049 \times 10^3$	0,99418
Leukosit	1,2	3,322	0,7262	0,09830
Trombosit	1,191	3,291	0,6050	0,004896
PO	3,453	31,60	0,6824	$4,18 \times 10^{-7}$
DK	18,65	$1,257 \times 10^8$	$1,222 \times 10^3$	0,98782
<i>Epiru</i>	-0,4731	0,6231	$1,647 \times 10^3$	0,99977
<i>Cyclo</i>	4,659	$1,055 \times 10^2$	$1,434 \times 10^3$	0,99741
<i>Likelihood ratio</i>				0,02

Berdasarkan Tabel 4.30 memperlihatkan hasil estimasi parameter sehingga didapatkan persamaan Regresi *Cox Proportional Hazard* sebagai berikut

$$\begin{aligned}
 h(t) &= h_0(t) \exp(X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10}) \\
 &= h_0(t) \exp(-2,567X_1 - 3,196X_2 + 0,8062X_3 - 14,95X_4 + 1,2X_5 \\
 &\quad + 1,191X_6 + 3,453X_7 + 18,65X_8 - 0,4731X_9 + 4,659X_{10})
 \end{aligned}$$

Pengujian parameter yang dilakukan adalah uji serentak dan uji parsial (*Wald*). Uji serentak dimaksudkan untuk mengetahui variabel bebas yang berpengaruh signifikan terhadap variabel terikat dan pengujian dilakukan secara bersamaan. Variabel bebas berpengaruh terhadap variabel terikat didapatkan dengan menolak hipotesis H_0 apabila nilai *p-value Likelihood ratio* $< \alpha = 0,05$. Hipotesis yang digunakan dalam pengujian serentak yaitu

H_0 : variabel bebas tidak berpengaruh signifikan terhadap variabel respon

H_1 : minimal ada satu variabel bebas berpengaruh terhadap variabel respon

Nilai *p-value Likelihood ratio* pada Tabel 4.30 yaitu sebesar 0,02. Nilai tersebut lebih kecil dari nilai signifikansi $\alpha = 0,05$ sehingga didapatkan keputusan hipotesis H_0 ditolak. Hal ini berarti bahwa terdapat minimal satu variabel bebas yang berpengaruh signifikan terhadap variabel responnya. Dengan demikian, model *Cox Proportional Hazard* yang terbentuk secara bersama-sama telah berpengaruh signifikan.

Uji parsial didapatkan dengan menolak hipotesis H_0 apabila nilai $p\text{-value} < \alpha = 0,05$. Hipotesis yang digunakan dalam pengujian parsial yaitu

H_0 : variabel bebas tidak berpengaruh signifikan terhadap variabel respon

H_1 : variabel bebas berpengaruh signifikan terhadap variabel respon

Berdasarkan Tabel 4.30 terdapat variabel yang signifikan yakni usia, IMT, trombosit dan PO. Hal ini dibuktikan dengan masing-masing nilai $p\text{-value}$ lebih kecil dari $\alpha = 0,05$ yakni sebesar $2,63 \times 10^{-6}$, $0,00509$, $0,04896$ dan $4,18 \times 10^{-7}$.

4.1.5 Analisis Survival Menggunakan Metode *Random Survival Forest*

Dalam skripsi ini, metode RSF digunakan untuk menganalisis data ketahanan hidup pasien penderita kanker payudara. Pengolahan data dilakukan dengan menggunakan *software* R dengan package *randomForestSRC*. Berikut ini merupakan hasil *output* dari pengolahan data dengan metode RSF:

Tabel 4.31 Hasil *Output* Pengolahan Data dengan Metode RSF

```

Sample size: 48
Number of deaths: 5
Number of trees: 1000
Forest terminal node size: 15
Average no. of terminal nodes: 1.994
No. of variables tried at each split: 4
Total no. of variables: 10
Resampling used to grow trees: swor
Resample size used to grow trees: 30
Analysis: RSF
Family: surv
Splitting rule: logrank
*random*
Number of random split points: 3
(OOB) CRPS: 0.15495516
(OOB) Requested performance error: 0.15333333

```

Berdasarkan hasil *output* dari pengolahan data dengan metode RSF, dapat dilihat bahwa *sample size* menunjukkan banyaknya sampel yang digunakan sebagai data *bootstrap* yaitu 70% dari 68 data pasien kanker payudara. *Number of death* atau jumlah kematian sebanyak 5, hal ini menyatakan bahwa terdapat 5 subjek yang meninggal. *Number of trees* menunjukkan jumlah pohon yang dibangun untuk menyusun *forest* yaitu sebanyak 1000 pohon. *Forest terminal node size* adalah jumlah sampel *training* yang ada pada node terminal. Artinya, pohon *survival* dibangun sampai node terminal memiliki ukuran 15 atau dengan kata lain pohon *survival* tidak dapat tumbuh lagi pada ukuran node terminal yang kurang dari 15 node. Sedangkan average no. of terminal nodes menunjukkan jumlah rata-rata pada terminal node adalah 1,994.

Selain itu, *no. of variables tried at each split* menunjukkan banyaknya variabel kandidat dalam pemisahan node yang diperoleh dari akar jumlah variabel bebas. Oleh karena variabel bebas pada penelitian sebanyak 4 variabel, maka jumlah variabel kandidat untuk pemisahan node adalah $\sqrt{10}$ variabel. *Mtry* (banyaknya variabel kandidat dalam pemisahan node) dengan aturan tersebut memberikan hasil yang optimal.

Sementara itu, pengambilan sampel ulang untuk *sample bootstrap* (resampling) dilakukan dengan pengembalian atau *swor*. Kemudian “*analysis : RSF* dan *family : surv*” menunjukkan bahwa pengolahan data dilakukan menggunakan metode *random survival forest* yang merupakan salah satu metode dalam analisis *survival*. *Splitting rule* atau aturan *splitting* yang digunakan dalam analisis ini adalah aturan *logrank random*. *Number of random split points* menunjukkan jumlah titik pemisahan acaknya adalah 3. Titik pemisahan dapat dipilih secara acak. Nilai titik pemisahan yang kecil diperlukan dalam pengaturan yang melibatkan campuran variabel diskrit dan kontinu untuk menghindari pemisahan bias terhadap variabel kontinu. Dari tabel 4.31 juga terlihat bahwa hasil kesalahan prediksi yang diperoleh adalah sebesar 15,3%.

Tujuan utama dalam analisis *survival* menggunakan metode RSF ini adalah untuk mendapatkan variabel penting yang mampu meningkatkan kinerja prediksi. Pada penelitian ini, pemilihan variabel penting dilakukan untuk mengetahui

variabel mana saja yang dapat memprediksi ketahanan hidup pasien penderita kanker payudara dengan baik. Variabel dengan nilai kepentingan (*importance*) yang positif menunjukkan bahwa variabel tersebut memiliki kemampuan prediksi yang baik. Sementara itu, jika nilai kepentingannya nol atau negatif maka variabel tersebut nonprediktif. Pada Tabel 4.32 memperlihatkan nilai *importance* untuk masing-masing variabel bebas.

Tabel 4.32 Nilai *Importance* Variabel Bebas

Variabel	<i>Importance</i>	<i>Relative Imp</i>
PO	0,4792	1,0000
DK	0,1047	0,2185
Leukosit	0,0535	0,1115
RPL	0,0129	0,0270
Cyclo	0,0072	0,0150
Epiru	0,0051	0,0106
HB	0,0033	0,0068
IMT	-0,0012	-0,0025
Usia	-0,0062	-0,0129
Trombosit	-0,0192	-0,0400

Setiap variabel yang memiliki nilai *importance* positif merupakan variabel yang prediktif atau dengan kata lain variabel tersebut dapat memprediksi ketahanan hidup pasien penderita kanker payudara secara signifikan, begitu juga sebaliknya. Berdasarkan Tabel 4.31, terlihat bahwa nilai *importance* pada variabel PO, DK, leukosit, RPL, cyclo, epiru dan HB adalah bernilai positif sehingga variabel-variabel tersebut dapat memprediksi ketahanan hidup pasien penderita kanker payudara secara signifikan. Nilai VIMP untuk variabel PO adalah 0,4792, artinya variabel PO memberikan peningkatan 47,92% dalam ketahanan hidup pasien penderita kanker payudara. Nilai VIMP untuk variabel DK adalah 0,1047, artinya variabel DK memberikan peningkatan sebesar 10,47% dalam ketahanan hidup pasien penderita kanker payudara. Nilai VIMP untuk variabel leukosit adalah 0,0535, artinya variabel leukosit memberikan peningkatan sebesar 5,35% dalam

ketahanan hidup pasien penderita kanker payudara. Nilai VIMP untuk variabel RPL adalah 0,0129, artinya variabel RPL memberikan peningkatan sebesar 1,29% dalam ketahanan hidup pasien penderita kanker payudara. Nilai VIMP untuk variabel *cyclo* adalah 0,0072, artinya variabel *cyclo* memberikan peningkatan sebesar 0,72% dalam ketahanan hidup pasien penderita kanker payudara. Nilai VIMP untuk variabel *epiru* adalah 0,0051, artinya variabel *epiru* memberikan peningkatan sebesar 0,51% dalam ketahanan hidup pasien penderita kanker payudara. Nilai VIMP variabel HB dengan nilai VIMP adalah 0,0033 yang artinya variabel memberikan peningkatan sebesar 0,33%.

Jika dilihat dari nilai *importancenya*, variabel PO merupakan variabel dengan nilai *importance* yang lebih besar diantara variabel lainnya, artinya variabel PO merupakan variabel yang paling prediktif dibandingkan dengan variabel lain. Sementara itu, variabel HB merupakan variabel yang cukup prediktif namun dinilai tidak terlalu meningkatkan kinerja prediksi karena nilai *importancenya* yang cukup kecil dan hampir mendekati nol dibandingkan dengan variabel lainnya. Sedangkan variabel IMT, usia dan trombosit nilai *importancenya* negatif sehingga variabel-variabel tersebut dinilai tidak prediktif. Sedangkan pada variabel IMT, usia dan trombosit dengan masing-masing nilai *importancenya* yaitu -0,0012, -0,0062 dan -0,0192, artinya variabel tersebut tidak prediktif karena tidak mampu meningkatkan kinerja prediksi yang ditandai dengan nilai yang negatif.

4.2 Pembahasan

Penelitian ini berkaitan dengan perbandingan dua metode yaitu *Cox Proportional Hazard* dan *Random Survival Forest* untuk mengetahui ketahanan hidup pasien penderita kanker payudara. Cara membandingkan kedua model yaitu dengan melihat nilai *C-index* dimana model yang memiliki nilai *C-index* terkecil merupakan model terbaik. Pada penelitian ini, data yang digunakan yaitu data sekunder pasien penderita kanker payudara yang dirawat di RSD dr. Soebandi Jember dengan periode penelitian selama Januari 2022 sampai dengan Januari 2023. Setelah dilakukan penelitian, didapatkan 68 data pasien penderita kanker payudara. Pada Tabel 4.1 dapat diketahui bahwa data *event* terbagi menjadi dua

yaitu data *event* tersensor dan tidak tersensor (teramati) dengan masing-masing data sebanyak 63 pasien dan 5 pasien.. Data dengan *event* tidak tersensor adalah meninggalnya pasien penderita kanker payudara saat masih menjalani pengobatan. Sedangkan data dengan *event* tersensor adalah pasien penderita kanker payudara yang tidak mengalami *event* (meninggal) selama waktu yang telah ditentukan seperti pasien pulang dengan keadaan masih hidup, berobat jalan (rawat jalan), menolak untuk dirawat dan dirujuk ke rumah sakit lain. Variabel bebas yang diduga memengaruhi ketahanan hidup pasien penderita kanker payudara pada penelitian ini adalah usia, indeks massa tubuh (IMT), Riwayat penyakit lain (RPL), hemoglobin (HB), leukosit, trombosit, penggunaan obat (PO), disiplin kemoterapi (DK), penggunaan dosis *epirubicin* dan penggunaan dosis *cyclospamid*.

Hasil analisis deskriptif dan nilai *hazard* setiap variabel didapatkan proporsi yang paling tinggi mengalami *event* (meninggal) yaitu variabel dengan kategori $40 \leq \text{usia} < 50$, $25,1 \leq \text{IMT} < 30$, ada riwayat penyakit lain, HB rendah, leukosit rendah, trombosit rendah, obat diganti, tidak mengikuti kemoterapi, tidak menggunakan dosis *epirubicin* dan tidak menggunakan dosis *cyclospamid*. Selanjutnya variabel bebas dilakukan uji asumsi *proportional hazard* dengan menggunakan *Goodness of Fit* yaitu *Schoenfeld Residuals*. Tabel 4.2 menunjukkan bahwa semua variabel memenuhi asumsi *proportional hazard*. Variabel yang telah memenuhi asumsi *proportional hazard* dilakukan estimasi parameter. Berdasarkan hasil estimasi parameter *Cox Proportional Hazard* diperoleh model terbaik pada Tabel 4.30 dan didapatkan persamaan sebagai berikut

$$h(t) = h_0(t) \exp (2,63 \times 10^{-6} X_1 + 0,00509 X_2 + 0,04896 X_6 + 4,18 \times 10^{-7} X_7)$$

dari model tersebut, dapat dilihat bahwa variabel yang signifikan berpengaruh terhadap ketahanan hidup pasien penderita kanker payudara adalah usia, indeks massa tubuh, trombosit dan penggunaan obat. Sedangkan pada *Random Survival Forest*, variabel yang signifikan berpengaruh adalah penggunaan obat, disiplin kemoterapi, leukosit, Riwayat penyakit lain, penggunaan dosis *cyclospamid*, penggunaan dosis *epirubicin* dan hemoglobin. Model *Cox Proportional Hazard* dan *Random Survival Forest* menunjukkan variabel penggunaan obat merupakan variabel yang signifikan berpengaruh ketahanan hidup pasien penderita kanker

payudara pada kedua model. Penggunaan obat yang berpengaruh pada pasien penderita kanker payudara yaitu penggunaan obat tetap dengan dosis tetap dan obat diganti.

Perbandingan kedua metode dalam penelitian ini dilakukan dengan menggunakan metode *Harrell's Concordance Index* atau yang biasa dikenal dengan *C-Index*. Membandingkan kedua metode dilakukan untuk melihat manakah metode yang cocok diterapkan pada data. Membandingkan metode dengan *C-Index* dilakukan dengan melihat nilai *error C-Indexnya*, semakin kecil nilai *error* yang dihasilkan maka semakin baik kinerja metode yang digunakan. Dengan bantuan *software R*, diperoleh nilai *error C-Index* adalah

Tabel 4.33 *Harrell's Concordance Index* Untuk Perbandingan Metode

	CPH	RSF
<i>C-Index</i>	0,97	0,15

Berdasarkan Tabel 4.32, dapat dilihat bahwa nilai *error C-index* yang dihasilkan oleh metode RSF adalah 15%, dimana nilai tersebut lebih kecil dari metode CPH yaitu 97% sehingga dapat disimpulkan bahwa metode yang lebih sesuai untuk data ketahanan hidup pasien penderita kanker payudara adalah metode *Random Survival Forest* (RSF). Hal tersebut relevan dengan kinerja metode RSF yang memang cocok digunakan untuk data berdimensi besar dan memiliki kovariat banyak. Oleh karena itu, RSF dapat meningkatkan akurasi secara signifikan dibandingkan CPH, terutama untuk data dengan banyak variabel kontinu yang memiliki efek non linier pada hasil prediksi.

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan diperoleh kesimpulan yaitu faktor-faktor yang memengaruhi ketahanan hidup pasien penderita kanker payudara pada model *Cox Proportional Hazard* yaitu usia, indeks massa tubuh, trombosit dan penggunaan obat. Sedangkan pada model *Random Survival Forest* faktor-faktor yang berpengaruh secara signifikan dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara yaitu penggunaan obat, disiplin kemoterapi, leukosit, riwayat penyakit lain, penggunaan dosis *cyclospamid*, penggunaan dosis *epirubicin* dan hemoglobin. Pada model *Cox Proportional Hazard* dan *Random Survival Forest* menunjukkan variabel penggunaan obat merupakan variabel yang signifikan berpengaruh terhadap ketahanan hidup pasien penderita kanker payudara. Penggunaan obat yang berpengaruh pada pasien penderita kanker payudara yaitu penggunaan obat tetap dengan dosis tetap dan obat diganti. Namun, berdasarkan hasil perbandingan model *Cox Proportional Hazard* dan *Random Survival Forest* dalam memprediksi waktu ketahanan hidup pasien penderita kanker payudara menunjukkan bahwa metode RSF lebih baik daripada CPH. Hal ini dikarenakan RSF memiliki nilai *C-Index* lebih kecil yaitu sebesar 0,15 sedangkan nilai *C-Index* pada CPH sebesar 0,97.

5.2 Saran

Penelitian ini membandingkan antara model *cox proportional hazard* dan *random survival forest*, untuk penelitian selanjutnya diharapkan dapat menggunakan model *survival* lainnya untuk mengetahui manakah metode yang paling baik dalam menyeleksi variabel prediktif. Selain itu, pada penelitian ini menggunakan nilai *C-index* terkecil sebagai pemilihan model terbaik, sehingga untuk penelitian selanjutnya dapat menggunakan metode lain dalam menentukan pemilihan model terbaik. Penelitian selanjutnya juga diharapkan dapat menggunakan objek penelitian selain dalam bidang kesehatan dan melakukan analisis dengan data yang lebih banyak.

DAFTAR PUSTAKA

- Chmielewska, A.P. dan A. Matuszyk. 2020. Application Of The Random Survival Forests Method In The Bankruptcy Prediction For Small and Medium Enterprises. *Warsaw School of Economics* 1(44): 1-16.
- Departemen Kesehatan Republik Indonesia. 2016. *Profil Kesehatan Republik Indonesia Tahun 2015*. Jakarta: Dinas Kesehatan RI.
- Dietz, A. T., H. D. Nelson., B. Zakher., A. Cantor., R. Fu., J. Griffin., E. S. O'Meara., D. S. M. Buist., K. Kerlikowske., N. T. V. Ravesteyn., J. S. Mandelblatt, dan D. L. Miglioretti. 2012. Risk Factors for Breast Cancer for Women Aged 40 to 49 Years: A Systematic Review and Meta-Analysis. *Annals of Internal Medicine*. 156(9):635-648.
- Ghodratollah, R., M. Safari., J. Faradmal., M. Abbasi, dan S. Khazaei. 2020. Factor Affecting The Survival of Patients With Colorectal Cancer Using Random Survival Forest. *Journal of Gastrointestinal*. 20(3):1-8.
- GLOBOCAN. 2020. International gency for Research on Cancer: Press Release Latest Global Cancer Data 2018 Retrieved 29 September 2019 from World Health Organization. <https://www.who.int/cancer/PRGlobocanFinal.pdf?ua=1>. [Diakses pada 28 Oktober 2022].
- Harlan, J. 2017. *Analisis Survival*. Jakarta: Gunadarma.
- Ishwaran, H., U. B. Kogalur., E. H. Blackstone, dan Michael S. Lauer. 2008. Random Survival Forest. *The Annals of Applied Statistics*. 2:841-860.
- Kaminska, M., T. Ciszewski, dan K. Szatan. 2015. *Breast Cancer Risk Factors*. Poland: Prz Menopauzalny.
- Kementrian Kesehatan RI. 2016a. *Profil Kesehatan Indonesia Tahun 2015*. Jakarta: Kemenkes RI.
- Kementerian Kesehatan RI. 2016b. *Pusat Data dan Informasi Kementerian Kesehatan RI Bulan Peduli Kanker Payudara*. Jakarta: INFODATIN.
- Kleinbaum, D. G. dan M. Klein. 2012. *Survival Analysis A Self-Learning Text Third Edition*. New York: Spinger Science Business Media.
- Qomaria, T., M. Fatekurohman, dan D. Anggraeni. 2019. Aplikasi Model Cox Proportional Hazard pada Pasien Stroke RSUD Balung Kabupaten Jember. *Indonesian Journal of Applied Statistic*. 2(2): 94-112.

Lee, E.T. dan J. W. Wang. 2003. *Statistical Methods for Survival Data Analysis Third Edition*. New Jersey: John Wiley and Sons.

Nurchahyo, J. 2010. *Awas Bahaya Kanker Rahim dan Kanker Payudara*. Yogyakarta: Wahana Totalita Publisher.

Rebolj, M., J. Bonde., D. Ejegod., S. Preisler., C. Rygaard, dan E. Lynge. 2015. A daunting challenge: Human Papillomavirus assays and cytology in primary cervical screening of women below age 30 years. *European Journal of Cancer*. 51(11):1456-1466.

Suryaningsih, E. K. dan B. E. Sukaca. 2009. *Kupas Tuntas Kanker Payudara*. Yogyakarta: Paradigma Indonesia.

Wulandari, R. 2012. Peran Kemoterapi Eksterna Terhadap Penderita Kanker Payudara Stadium Lokal-Lanjut. *Skripsi*. Semarang: Fakultas Kedokteran Universitas Diponegoro.

Xianxin, Q., J. Gao., J. Yang., J. Hu., W. Hu., L. Kong, dan J. J. Lu. 2020. A Comparison Study of Machine Learning (Random Survival Forest) and Classic Statistic (Cox Proportional Hazards) for Predicting Progression in High-Grade Glioma after Proton and Carbon Ion Radiotherapy. *Frontiers in Oncology*. 10: 1-10.

LAMPIRAN

Lampiran 1. Data Pasien Kanker Payudara

No.	Usia	IMT	RPL	HB	Leukosit	Trombosit	PO	DK	Epiru	Waktu	Status
1	4	2	0	1	1	1	0	0	0	97	0
2	0	3	0	0	1	1	0	0	0	78	0
3	0	1	0	1	1	0	0	0	0	65	0
4	2	1	1	0	2	2	0	1	0	44	0
5	2	1	0	0	1	2	0	0	0	45	0
6	1	1	1	0	1	0	1	1	0	67	0
7	1	1	1	0	0	0	2	1	0	28	1
8	3	1	0	1	1	1	0	1	0	31	0
9	1	2	0	0	0	2	0	0	1	177	0
10	0	2	1	1	1	1	0	0	1	135	0
11	3	1	0	0	2	1	0	1	0	141	0
12	1	3	0	1	1	1	0	1	0	90	0
13	1	1	0	0	0	0	2	1	0	55	1
14	1	0	0	2	1	1	0	0	1	75	0
15	2	1	0	1	1	1	1	1	0	65	0
16	1	2	0	1	1	1	0	0	1	88	0
17	1	1	0	0	1	1	0	0	0	90	0
18	1	2	0	0	1	2	0	0	0	23	0
19	2	1	1	0	2	0	1	1	0	45	0
20	3	2	0	0	2	2	1	0	0	23	0
∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴	∴
60	4	3	0	0	2	2	0	0	1	75	0
61	2	3	0	0	0	1	0	0	0	84	0
62	1	2	0	0	2	0	0	1	0	56	0
63	2	3	0	0	2	1	0	1	0	76	0
64	3	1	1	0	1	2	2	1	0	79	0
65	2	1	1	0	2	1	0	0	0	87	0
66	2	2	0	1	1	1	0	0	0	45	0
67	2	1	1	0	2	2	0	1	0	67	1
68	3	1	0	0	2	1	0	0	1	73	0

Lampiran 2. Pemanggilan data

```
#Data Kanker Payudara
#Memanggil package yang diperlukan
library(survival)
library(readxl)
#Memanggil data
DATAKU<-read.csv(file.choose(), header=TRUE)
View(DATAKU)
```

Lampiran 3. Script dan Output Program R Analisis Deskriptif

```
> #Analisis Deskriptif
> table(DATAKU$Status,DATAKU$Usia)
  0  1  2  3  4
  0  7 15 22 16  3
  1  0  3  1  1  0
> tapply(DATAKU$Status,DATAKU$Usia,summary,na.rm=T);
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  0      0      0      0      0      0
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.0000 0.0000 0.0000 0.1667 0.0000 1.0000
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.00000 0.00000 0.00000 0.04348 0.00000 1.00000
$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.00000 0.00000 0.00000 0.05882 0.00000 1.00000
$`4`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  0      0      0      0      0      0
```

```

> table(DATAKU$Status,DATAKU$IMT)
  0 1 2 3
0 5 35 13 10
1 0 3 2 0
> tapply(DATAKU$Status,DATAKU$IMT,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.000000 0.000000 0.000000 0.07895 0.000000 1.000000
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 0.0000  0.0000  0.0000  0.1333  0.0000  1.0000
$`3`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000

> table(DATAKU$Status,DATAKU$RPL)
  0 1
0 40 23
1 1 4
> tapply(DATAKU$Status,DATAKU$RPL,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
0.000000 0.000000 0.000000 0.02439 0.000000 1.000000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 0.0000  0.0000  0.0000  0.1481  0.0000  1.0000

```



```

> table(DATAKU$Status,DATAKU$HB)
  0 1 2
0 37 25 1
1 5 0 0
> tapply(DATAKU$Status,DATAKU$HB,summary,na.rm=T);
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
0.000 0.000 0.000 0.119 0.000 1.000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0      0      0      0      0      0
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0      0      0      0      0      0

> table(DATAKU$Status,DATAKU$Leukosit)
  0 1 2
0 4 33 26
1 2 0 3
> tapply(DATAKU$Status,DATAKU$Leukosit,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
0.0000 0.0000 0.0000 0.3333 0.7500 1.0000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  0      0      0      0      0      0
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
0.0000 0.0000 0.0000 0.1034 0.0000 1.0000

> table(DATAKU$Status,DATAKU$Trombosit)

```

```

0 1 2
0 9 37 17
1 3 0 2
> tapply(DATAKU$Status,DATAKU$Trombosit,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00  0.00  0.00  0.25  0.25  1.00
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0      0      0      0      0      0
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000 0.0000 0.0000 0.1053 0.0000 1.0000

> table(DATAKU$Status,DATAKU$PO)
 0 1 2
0 54 6 3
1 2 0 3
> tapply(DATAKU$Status,DATAKU$PO,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.00000 0.00000 0.03571 0.00000 1.00000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0      0      0      0      0      0
$`2`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0   0.0   0.5   0.5   1.0   1.0

> table(DATAKU$Status,DATAKU$DK)
 0 1

```

```

0 33 30
1 0 5
> tapply(DATAKU$Status,DATAKU$DK,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0      0      0      0      0      0
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000 0.0000 0.0000 0.1429 0.0000 1.0000

> table(DATAKU$Status,DATAKU$Epiru)
 0 1
0 53 10
1 5 0
> tapply(DATAKU$Status,DATAKU$Epiru,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00000 0.00000 0.00000 0.08621 0.00000 1.00000
$`1`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0      0      0      0      0      0

> table(DATAKU$Status,DATAKU$Cyclo)
 0 1 2
0 35 25 3
1 5 0 0
> tapply(DATAKU$Status,DATAKU$Cyclo,summary,na.rm=T)
$`0`
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000 0.000 0.000 0.125 0.000 1.000
$`1`

```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0 0 0 0 0 0
```

\$`2`

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
0 0 0 0 0 0
```

Lampiran 4. Script dan Output Fungsi Survival

```
> #Fungsi Survival
> Survfit1<-survfit(Surv(Waktu,Status)~Usia,data=DATAKU)
> summary(Survfit1) #output lebih lengkap
Call: survfit(formula = Surv(Waktu, Status) ~ Usia, data = DATAKU)
```

Usia=0

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

Usia=1

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
6 18 1 0.944 0.0540 0.844 1
28 16 1 0.885 0.0763 0.748 1
55 10 1 0.797 0.1085 0.610 1
```

Usia=2

```
time n.risk n.event survival std.err lower 95% CI
```

```
67.0000 12.0000 1.0000 0.9167 0.0798 0.7729
```

upper 95% CI

1.0000

Usia=3

```
time n.risk n.event survival std.err lower 95% CI
```

```
157 1 1 0 NaN NA
```

upper 95% CI

NA

Usia=4

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

> Survfit2<-survfit(Surv(Waktu,Status)~IMT,data=DATAKU)

> summary(Survfit2)

Call: survfit(formula = Surv(Waktu, Status) ~ IMT, data = DATAKU)

IMT=0

time n.risk n.event survival std.err lower 95% CI upper 95% CI

IMT=1

time n.risk n.event survival std.err lower 95% CI upper 95% CI

28 37 1 0.973 0.0267 0.922 1

55 26 1 0.936 0.0448 0.852 1

67 18 1 0.884 0.0659 0.763 1

IMT=2

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 15 1 0.933 0.0644 0.815 1

157 2 1 0.467 0.3316 0.116 1

IMT=3

time n.risk n.event survival std.err lower 95% CI upper 95% CI

> Survfit3<-survfit(Surv(Waktu,Status)~RPL,data=DATAKU)

> summary(Survfit3)

Call: survfit(formula = Surv(Waktu, Status) ~ RPL, data = DATAKU)

RPL=0

time n.risk n.event survival std.err lower 95% CI

55.0000 32.0000 1.0000 0.9688 0.0308 0.9103

upper 95% CI

1.0000

RPL=1

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6 27 1 0.963 0.0363 0.894 1

28 25 1 0.924 0.0514 0.829 1

67 13 1 0.853 0.0832 0.705 1

```
157 1 1 0.000 NaN NA NA
```

```
> Survfit4<-survfit(Surv(Waktu,Status)~HB,data=DATAKU)
```

```
> summary(Survfit4)
```

```
Call: survfit(formula = Surv(Waktu, Status) ~ HB, data = DATAKU)
```

```
HB=0
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	42	1	0.976	0.0235	0.931	1
28	39	1	0.951	0.0337	0.887	1
55	31	1	0.920	0.0444	0.837	1
67	24	1	0.882	0.0568	0.778	1
157	2	1	0.441	0.3132	0.110	1

```
HB=1
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	42	1	0.976	0.0235	0.931	1

```
HB=2
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	42	1	0.976	0.0235	0.931	1

```
> Survfit5<-survfit(Surv(Waktu,Status)~Leukosit,data=DATAKU)
```

```
> summary(Survfit5)
```

```
Call: survfit(formula = Surv(Waktu, Status) ~ Leukosit, data = DATAKU)
```

```
Leukosit=0
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
28	6	1	0.833	0.152	0.583	1
55	4	1	0.625	0.213	0.320	1

```
Leukosit=1
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	42	1	0.976	0.0235	0.931	1

```
Leukosit=2
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	29	1	0.966	0.0339	0.901	1
67	14	1	0.897	0.0735	0.763	1
157	1	1	0.000	NaN	NA	NA

> Survfit6<-survfit(Surv(Waktu,Status)~Trombosit,data=DATAKU)

> summary(Survfit6)

Call: survfit(formula = Surv(Waktu, Status) ~ Trombosit, data = DATAKU)

Trombosit=0

time n.risk n.event survival std.err lower 95% CI upper 95% CI

28	12	1	0.917	0.0798	0.773	1
55	8	1	0.802	0.1279	0.587	1
157	1	1	0.000	NaN	NA	NA

Trombosit=1

time n.risk n.event survival std.err lower 95% CI upper 95% CI

Trombosit=2

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6	19	1	0.947	0.0512	0.852	1
67	8	1	0.829	0.1195	0.625	1

> Survfit7<-survfit(Surv(Waktu,Status)~PO,data=DATAKU)

> summary(Survfit7)

Call: survfit(formula = Surv(Waktu, Status) ~ PO, data = DATAKU)

PO=0

time n.risk n.event survival std.err lower 95% CI upper 95% CI

67	33	1	0.970	0.0298	0.913	1
157	2	1	0.485	0.3432	0.121	1

PO=1

time n.risk n.event survival std.err lower 95% CI upper 95% CI

PO=2

time n.risk n.event survival std.err lower 95% CI upper 95% CI

6	6	1	0.833	0.152	0.583	1
28	5	1	0.667	0.192	0.379	1
55	3	1	0.444	0.222	0.167	1

```
> Survfit8<-survfit(Surv(Waktu,Status)~DK,data=DATAKU)
```

```
> summary(Survfit8)
```

```
Call: survfit(formula = Surv(Waktu, Status) ~ DK, data = DATAKU)
```

```
DK=0
```

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
DK=1
```

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	35	1	0.971	0.0282	0.918	1
28	34	1	0.943	0.0392	0.869	1
55	23	1	0.902	0.0549	0.800	1
67	17	1	0.849	0.0729	0.717	1
157	1	1	0.000	NaN	NA	NA

```
> Survfit9<-survfit(Surv(Waktu,Status)~Epiru,data=DATAKU)
```

```
> summary(Survfit9)
```

```
Call: survfit(formula = Surv(Waktu, Status) ~ Epiru, data = DATAKU)
```

```
Epiru=0
```

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	58	1	0.983	0.0171	0.950	1
28	54	1	0.965	0.0246	0.917	1
55	38	1	0.939	0.0347	0.874	1
67	27	1	0.904	0.0478	0.815	1
157	1	1	0.000	NaN	NA	NA

```
Epiru=1
```

```
time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
> Survfit10<-survfit(Surv(Waktu,Status)~Cyclo,data=DATAKU)
```

```
> summary(Survfit10)
```

```
Call: survfit(formula = Surv(Waktu, Status) ~ Cyclo, data = DATAKU)
```

```
Cyclo=0
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	40	1	0.975	0.0247	0.928	1
28	38	1	0.949	0.0349	0.883	1
55	27	1	0.914	0.0482	0.824	1
67	21	1	0.871	0.0625	0.756	1
157	2	1	0.435	0.3094	0.108	1

Cyclo=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
------	--------	---------	----------	---------	--------------	--------------

Cyclo=2

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
------	--------	---------	----------	---------	--------------	--------------

Lampiran 5. Script dan Output Uji Asumsi Cox Proportional Hazard

```
coxph1<-coxph(Surv(Waktu,Status)~ Usia + IMT + RPL + HB + Leukosit + Trombosit + PO + DK + Epiru + Cyclo, data=DATAKU)
```

```
> cox.zph(coxph1)
```

	chisq	df	p
Usia	1.50e-01	1	0.70
IMT	6.67e-01	1	0.41
RPL	2.10e-01	1	0.65
HB	4.38e-07	1	1.00
Leukosit	3.88e-01	1	0.53
Trombosit	4.67e-02	1	0.83
PO	1.75e+00	1	0.19
DK	3.83e-06	1	1.00
Epiru	5.42e-04	1	0.98
Cyclo	1.16e-06	1	1.00
GLOBAL	1.43e+01	10	0.16

Lampiran 6. Script dan Output Uji Estimasi Parameter Cox Proportional Hazard

```
#ESTIMASI PARAMETER
```

```
> cox<-
coxph(Surv(Waktu,Status==1)~Usia+IMT+RPL+HB+Leukosit+Trombosit+PO+
DK+Epiru+Cyclo,data=DATAKU)
```

Warning message:

In coxph.fit(X, Y, istrat, offset, init, control, weights = weights, :

Ran out of iterations and did not converge

```
> summary(cox)
```

Call:

```
coxph(formula = Surv(Waktu, Status == 1) ~ Usia + IMT + RPL +
      HB + Leukosit + Trombosit + PO + DK + Epiru + Cyclo, data = DATAKU)
```

n= 68, number of events= 5

	coef	exp(coef)	se(coef)	z	Pr(> z)
Usia	-2.567e+00	7.678e-02	5.464e-01	-4.698	2.63e-06 ***
IMT	-3.196e+00	4.091e-02	1.141e+00	-2.801	0.00509 **
RPL	8.062e-01	2.239e+00	1.280e+00	0.630	0.52886
HB	-1.495e+01	3.216e-07	2.049e+03	-0.007	0.99418
Leukosit	1.200e+00	3.322e+00	7.262e-01	1.653	0.09830 .
Trombosit	1.191e+00	3.291e+00	6.050e-01	1.969	0.04896 *
PO	3.453e+00	3.160e+01	6.824e-01	5.061	4.18e-07 ***
DK	1.865e+01	1.257e+08	1.222e+03	0.015	0.98782
Epiru	-4.731e-01	6.231e-01	1.647e+03	0.000	0.99977
Cyclo	4.659e+00	1.055e+02	1.434e+03	0.003	0.99741

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

exp(coef) exp(-coef) lower .95 upper .95

Usia	7.678e-02	1.303e+01	0.026311	0.2240
IMT	4.091e-02	2.444e+01	0.004371	0.3829
RPL	2.239e+00	4.465e-01	0.182139	27.5343
HB	3.216e-07	3.110e+06	0.000000	Inf

Leukosit	3.322e+00	3.010e-01	0.800267	13.7877
Trombosit	3.291e+00	3.039e-01	1.005421	10.7712
PO	3.160e+01	3.164e-02	8.296175	120.3707
DK	1.257e+08	7.957e-09	0.000000	Inf
Epiru	6.231e-01	1.605e+00	0.000000	Inf
Cyclo	1.055e+02	9.477e-03	0.000000	Inf

Concordance = 0.974 (se = 0.027)
 Likelihood ratio test = 20.64 on 10 df, p=0.02
 Wald test = 62.53 on 10 df, p=1e-09
 Score (logrank) test = 26.65 on 10 df, p=0.003

Lampiran 7. *Script dan Output Program R Random Survival Forest*

```
>library(randomForestSRC)
> library(survival)
> KP<-read.csv(file.choose(), header=TRUE)
>KP
> dim(KP)
> 68*0.7
[1] 47.6
> set.seed(1000)
> train<-sample(1:68,48,replace=FALSE)
> traindata<-KP[train,]
> testdata<-KP[-train,]
> fit<-rfsrc(Surv(Waktu,
Status)~Usia+IMT+RPL+HB+Leukosit+Trombosit+PO+DK+Epiru+Cyclo,
data=traindata, ntree = 1000, mtry = 4, nsplit = 3, importance = TRUE)
> fit
```

Sample size: 48

Number of deaths: 5

Number of trees: 1000
 Forest terminal node size: 15
 Average no. of terminal nodes: 1.994
 No. of variables tried at each split: 4
 Total no. of variables: 10
 Resampling used to grow trees: swor
 Resample size used to grow trees: 30
 Analysis: RSF
 Family: surv
 Splitting rule: logrank *random*
 Number of random split points: 3
 (OOB) CRPS: 0.15495516
 (OOB) Requested performance error: 0.15333333

>plot(fit)

	Importance	Relative Imp
PO	0.4792	1.0000
DK	0.1047	0.2185
Leukosit	0.0535	0.1115
RPL	0.0129	0.0270
Cyclo	0.0072	0.0150
Epiru	0.0051	0.0106
HB	0.0033	0.0068
IMT	-0.0012	-0.0025
Usia	-0.0062	-0.0129
Trombosit	-0.0192	-0.0400

Lampiran 8. Script dan Output C-Index

#CPH

> CindexCPH<-

```
coxph(Surv(Waktu,Status)~Usia+IMT+RPL+HB+Leukosit+Trombosit+PO+DK+
Epiru+Cyclo,data=DATAKU)
```



```
> summary(CindexCPH)$concordance[1]
  C
  0.9741784
#RSF
> get.cindex(fit$yvar[,1], fit$yvar[,2], fit$predicted.oob)
[1] 0.1533333
```

