



TESIS- SS 142501

## ADDITIVE SURVIVAL LEAST SQUARE SVM UNTUK ANALISIS DATA SURVIVAL

Studi Simulasi dan Studi Kasus pada Data Pasien Kanker Serviks  
di RSUD dr. Soetomo Surabaya

CHUSNUL KHOTIMAH  
NRP. 06211550010213

DOSEN PEMBIMBING :  
Santi Wulan Purnami, M.Si., Ph.D.  
Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.

PROGRAM MAGISTER  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA, KOMPUTASI, DAN SAINS DATA  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2018



TESIS- SS 142501

## ADDITIVE SURVIVAL LEAST SQUARE SVM FOR SURVIVAL DATA ANALYSIS

Simulation Study and Case Study on Cervical Cancer Data at  
RSUD dr. Soetomo Surabaya

CHUSNUL KHOTIMAH  
NRP. 06211550010213

SUPERVISORS :  
Santi Wulan Purnami, M.Si., Ph.D.  
Dr.rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.

PROGRAM OF MAGISTER  
STATISTICS DEPARTMENT  
FACULTY OF MATHEMATICS, COMPUTATION, AND DATA SCIENCE  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2018

**ADDITIVE SURVIVAL LEAST SQUARE SVM UNTUK  
ANALISIS DATA SURVIVAL**  
Studi Simulasi dan Studi Kasus pada Data Pasien Kanker Serviks  
di RSUD dr. Soetomo Surabaya

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Sains (M.Si)  
di  
Institut Teknologi Sepuluh Nopember

Oleh :  
**CHUSNUL KHOTIMAH**  
**NRP. 06211550010213**

Tanggal Ujian : 16 Januari 2018  
Periode Wisuda : Maret 2018

Disetujui Oleh :




1. Dr. Santi Wulan Purnami, M.Si., Ph.D.  
NIP. 19720923 199803 2 001

(Pembimbing 1)




2. Dr. rer. pol. Dedy Dwi Prastyo, S.Si., M.Si.  
NIP. 19831204 200812 1 002

(Pembimbing 2)



3. Dr. Ir. Setiawan, MS.  
NIP. 19601030 198701 1 001

(Penguji)




4. Dr. Puhadi, M.Sc.  
NIP. 19620204 198701 1 001

(Penguji)



Dekan  
Fakultas Matematika, Komputasi, dan Sains Data  
Institut Teknologi Sepuluh Nopember



Prof. Dr. Basuki Widodo, M.Sc  
NIP. 19650506 198903 1 002

**ADDITIVE SURVIVAL LEAST SQUARE SVM UNTUK  
ANALISIS DATA SURVIVAL**  
(Studi Simulasi dan Studi Kasus pada Data Pasien Kanker Serviks di  
RSUD dr. Soetomo Surabaya)

Nama Mahasiswa : Chusnul Khotimah  
NRP : 06211550010213  
Dosen Pembimbing : Santi Wulan Purnami, M.Si., Ph.D.  
: Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.

**ABSTRAK**

Analisis *survival* adalah metode yang digunakan untuk menganalisis waktu suatu subjek mulai dari awal pengamatan sampai terjadinya kejadian. Awalnya, model *survival* menggunakan pendekatan parametrik. Selanjutnya Cox mengusulkan modelnya yaitu *proporsional hazard* yang bersifat semiparametrik. Kelebihan dari Cox PHM adalah penggunaan fungsi Likelihood parsial yang terbukti sesuai dengan fungsi Likelihood. Kelemahan dari Cox PHM adalah adanya asumsi proporsional dan asumsi linearitas dalam kovariat. Pada penelitian ini akan digunakan analisis *survival* dengan pendekatan nonparametrik yaitu *Additive Survival Least Square SVM (A-SURLSSVM)* yang dibandingkan dengan Cox PHM. Data yang digunakan merupakan data simulasi dan data riil. Data simulasi dibangkitkan dengan berbagai skenario meliputi jumlah data, persentase censoring, serta banyaknya variabel prediktor yang berbeda-beda baik untuk data dengan pola linier maupun nonlinier. Data riil yang digunakan merupakan data pasien kanker serviks di rumah sakit dr. Soetomo Surabaya. Kriteria kebaikan model yang digunakan dalam penelitian ini adalah Concordance indeks (C-index), Logrank dan *hazard ratio*. Semakin bagus suatu model jika nilai kriteria kebaikan model semakin tinggi. Pada penelitian ini didapatkan hasil bahwa performansi A-SURLSSVM lebih baik dari pada Cox PHM berdasarkan C-index dan Hazard Ratio baik pada data riil maupun simulasi. Selanjutnya, dilakukan *features selection* untuk mengetahui prediktor yang relevan mempengaruhi *survival time*. Dengan *features selection* pada data riil, dua dari dua belas variabel yang digunakan dikeluarkan dari model yaitu Komplikasi dan Riwayat KB sehingga dapat meningkatkan nilai C-index. Pada berbagai setting pada data simulasi performansi dari A-SURLSSVM selalu lebih baik dari Cox PHM. Pada data dengan pola nonlinier selisih performansi C-index antara A-SURLSSVM dan Cox PHM lebih besar dibandingkan pada data linier. Sehingga metode A-SURLSSVM cocok digunakan pada data nonlinier. Selain itu, C-index merupakan ukuran performansi yang paling bagus dibandingkan hazard ratio dan Logrank. Hasil *features selection* pada data simulasi sebanyak 100 replikasi menunjukkan bahwa interaksi antar variabel menghasilkan variabel *main confounder* dan *sub-main confounder*.

**Kata kunci:** A-SURLSSVM, C-indeks, *Features Selection*, *Hazard Ratio*, Kanker Serviks, Logrank, Prognostik Indeks, Studi Simulasi.

*(halaman ini sengaja dikosongkan)*

**ADDITIVE SURVIVAL LEAST SQUARE SVM FOR  
SURVIVAL DATA ANALYSIS**  
(Simulation Study and Case Study on Cervical Cancer Data at RSUD  
dr. Soetomo Surabaya)

Name of Student : Chusnul Khotimah  
ID : 06211550010213  
Supervisor : Santi Wulan Purnami, M.Si., Ph.D.  
Co-Supervisor : Dr. rer.pol. Dedy Dwi Prastyo, S.Si., M.Si.

**ABSTRACT**

Survival analysis is a method used to analyze the time of a subject from the beginning of observation until the occurrence of events. Initially, the survival model used a parametric approach. Furthermore, Cox proposed a model so Called Cox Proportional Hazard model (Cox-PHM) as a semiparametric approach. The parameter estimation in Cox PHM used partial Likelihood function proven in accordance with the Likelihood function. The weakness of the Cox PHM is the presence of proportional assumptions and linearity assumptions within the covariates. This research use survival analysis with nonparametric approach, that is Additive Survival Least Square SVM (A-SURLSSVM) and Cox PHM as a benchmark. The data used in this study are simulation and real datasets. The simulation datasets are generated with various scenarios including the sample sizes, the percentage of censoring, and the number of features for both linear and nonlinear data. The real dataset used is the cervical cancer patients in dr. Soetomo Surabaya. Performance measures use in this research are Concordance index (C-index), Logrank and hazard ratio. The better model if it produces higher performance measures. It is found that the performance of A-SURLSSVM is better than Cox PHM based on C-index and Hazard Ratio both on real and simulated datasets. Furthermore, feature selection is performed to determine which relevant predictors influence survival time. Using for the features selection of real datasets, two out of twelve variables are removed from the model, i.e Complications and Family History. At various settings in simulation data, the A-SURLSSVM is always better than Cox PHM. In the data with nonlinear pattern, the difference of C-index between A-SURLSSVM and Cox PHM is greater than in linear data. So the A-SURLSSVM method performs much better for nonlinear data. In addition, the C-index is the best performance measure compared to hazard ratio and logrank. The results of the features selection in the simulation data of with 100 replication show that the interaction between variables yields the main confounder and the sub-main confounder variables.

**Keywords:** A-SURLSSVM, C-index, Features Selection, Hazard Ratio, Cervical Cancer, Log Rank, Prognostic Index, Simulation Study.

*(halaman ini sengaja dikosongkan)*

## KATA PENGANTAR

Puji Syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik serta hidayah-Nya kepada penulis, sehingga penulis dapat menyelesaikan tesis yang berjudul **“Additif Survival Least Square SVM untuk Analisis Data Survival (Studi Simulasi dan Studi Kasus Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya)”**, ini dapat diselesaikan tepat pada waktunya.

Tesis ini diharapkan dapat memberikan manfaat kepada penulis pada khususnya dan pembaca pada umumnya. Penulisan tesis ini belum sempurna (tak ada gading yang tak retak), karena kesempurnaan hanya milik sang Pencipta, oleh karena itu penulisan mengharapkan kritik dan saran bagi pembaca untuk mendapatkan hasil yang lebih baik di kemudian hari.

Penulisan tesis ini tidak akan berjalan dengan lancar tanpa bantuan dan dukungan beberapa pihak, oleh karena itu pada kesempatan kali ini penulis ingin mengucapkan terimakasih kepada:

1. Kedua orang tua dan keluarga penulis yang telah memberikan dukungan baik secara moril dan materiil kepada penulis sehingga penulis dapat menyelesaikan dengan baik.
2. Bapak Dr. Suhartono, M.Sc., selaku Kepala Departemen Statistika FMKSD ITS.
3. Ibu Santi Wulan Purnami, M.Si., Ph.D., selaku pembimbing utama yang telah membimbing penulis hingga dengan penuh kesabaran, memberikan masukan, serta arahan demi terselesaikannya tesis ini.
4. Bapak Dr. rer. pol. Dedy Dwi Prasetyo, S.Si. M.Si., selaku pembimbing kedua yang telah memberikan saran, masukan serta arahan yang membangun demi kesempurnaan tesis ini.
5. Bapak Dr. Purhadi, M.Sc., selaku penguji yang memberikan masukan dan arahan pada penyelesaian tesis ini.
6. Bapak Dr. Setiawan, MS., selaku penguji yang memberikan masukan dan arahan pada penyelesaian tesis ini.



7. Seluruh Dosen Statistika ITS yang telah membekali ilmu kepada penulis sehingga dapat menyelesaikan tesis ini.
8. Lembaga Pengelola Dana Pendidikan (LPDP) yang telah memberikan kesempatan dan membiayai penulis untuk melanjutkan magister.
9. Kamu yang telah memberikan dukungan, terimakasih.
10. Teman-teman seperjuangan S2 Statistika ITS angkatan genap tahun 2015 (Halwa, Faroh, Amel, Titis, Riffani, Annisa, Ribka, Luthfi, Jaya, Rama, Samsul, dan Ismail) yang telah berjuang bersama-sama untuk menyelesaikan studi magister.
11. Semua teman, relasi, dan berbagai pihak yang tidak bisa penulis sebutkan namanya satu per satu yang telah membantu dalam penulisan laporan ini.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran diharapkan dari semua pihak untuk tahap pengembangan selanjutnya. Besar harapan penulis bahwa informasi sekecil apapun dalam tesis ini bermanfaat bagi semua pihak dan dapat menambah wawasan serta pengetahuan.

Surabaya, Januari 2018

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>LEMBAR PENGESAHAN</b> .....	iii
<b>ABSTRAK</b> .....	v
<b>ABSTRACT</b> .....	vii
<b>KATA PENGANTAR</b> .....	ix
<b>DAFTAR ISI</b> .....	xi
<b>DAFTAR GAMBAR</b> .....	xv
<b>DAFTAR TABEL</b> .....	xix
<b>DAFTAR LAMPIRAN</b> .....	xxi
<b>BAB 1 PENDAHULUAN</b> .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	5
1.3 Tujuan .....	6
1.4 Manfaat .....	6
1.5 Batasan Masalah .....	6
<b>BAB 2 TINJAUAN PUSTAKA</b> .....	7
2.1 Analisis <i>Survival</i> .....	7
2.2 Kurva <i>Survival</i> Kaplan Meier dan Uji Log Rank .....	15
2.3 <i>Cox Proportional Hazard</i> .....	16
2.3.1 Model <i>Proportional Hazard</i> .....	16
2.3.2 Estimasi Parameter Regresi <i>Cox PHM</i> .....	17
2.3.3 Pengujian Signifikansi Parameter .....	18
2.3.4 Asumsi <i>Proportional Hazard</i> .....	19
2.4 Support Vector Machine .....	22
2.5 Least Square Support Vector Machine .....	26
2.6 <i>Survival SVM</i> .....	28
2.7 <i>Additive Survival Least Square SVM</i> .....	30
2.8 Kriteria Kebaikan Model .....	32
2.9 Pembangkitan Data <i>Survival Time</i> .....	34

2.10 <i>Features Selection</i> dengan Menggunakan <i>Backward Elimination</i> ..	37
2.11 Kanker Serviks .....	37
2.11.1 Faktor Penyebab Kanker Serviks.....	39
2.11.2 Pencegahan Kanker Serviks .....	41
<b>BAB 3 METODOLOGI PENELITIAN</b> .....	43
3.1 Kajian Teoritis Metode A-SURLSSVM.....	43
3.2 Kerangka Konsep.....	43
3.3 Aplikasi Metode A-SURLSSVM .....	46
3.3.1 Sumber Data dan Variabel Penelitian .....	46
3.3.2 Struktur Data.....	50
3.3.3 Langkah Analisis .....	51
<b>BAB 4 ANALISIS DAN PEMBAHASAN</b> .....	55
4.1 Kajian Teoritis Persamaan Prognostik Indeks Model A-SURLSSVM .....	55
4.2 Penerapan Metode A-SURLSSVM dan Cox PHM pada Data Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya .....	61
4.2.1 Karakteristik Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya.....	61
4.2.2. Kurva Survival Kaplan Meier dan Uji Log rank .....	66
4.2.3 Analisis Survival dengan Cox PHM dan A-SURLSSVM .....	80
4.3 Penerapan Metode A-SURLSSVM pada Data Simulasi .....	87
4.3.1 Penerapan dengan Persentase Sensoring dan Jumlah Data yang Berbeda pada Data dengan Pola Linier .....	88
4.3.2 Penerapan dengan Persentase Sensoring dan Jumlah Data yang Berbeda pada Data dengan Pola Nonlinier .....	89
4.3.3 Penerapan dengan Persentase Sensoring dan Jumlah Variabel yang Berbeda pada Data dengan Pola Linier .....	90
4.3.4 Penerapan dengan Persentase Sensoring dan Jumlah Variabel yang Berbeda pada Data dengan Pola Nonlinier .....	92
4.3.5 Penerapan Metode dengan Persentase Sensoring yang Berbeda pada Data dengan Pola Linier dan Nonlinier .....	93

4.3.6 Penerapan dengan Jumlah Variabel yang Berbeda pada dengan Pola Data Linier dan Nonlinier .....	94
4.3.7 Penerapan Metode A-SURLSSVM dengan <i>Features Selection</i> pada Data Simulasi .....	96
<b>BAB 5 KESIMPULAN DAN SARAN</b> .....	99
5.1 Kesimpulan .....	99
5.2 Saran .....	100
<b>DAFTAR PUSTAKA</b> .....	101
<b>LAMPIRAN</b> .....	107

*(halaman ini sengaja dikosongkan)*

## DAFTAR GAMBAR

Gambar 2.1	Contoh Data Tersensor Tipe I .....	8
Gambar 2.2	Contoh Data Tersensor Tipe II .....	9
Gambar 2.3	Contoh Data Tersensor Tipe III .....	10
Gambar 2.4	Ilustrasi Data Sensor Kanan .....	10
Gambar 2.5	Ilustrasi Data Sensor Kiri .....	11
Gambar 2.6	Ilustrasi Data Sensor Interval .....	11
Gambar 2.7	Ilustrasi $S(t)$ secara Teori .....	12
Gambar 2.8	Ilustrasi $S(t)$ pada Penerapan.....	13
Gambar 2.9	Ilustrasi Kurva <i>Survival</i> Kaplan Meier .....	15
Gambar 2.10	Ilustrasi Kurva $-\ln(-\ln S(t))$ .....	20
Gambar 2.11	Ilustrasi Grafik <i>Observed versus Expected</i> .....	20
Gambar 2.12	Konsep <i>Hyperplane</i> pada SVM.....	23
Gambar 2.13	<i>Mapping</i> dari Dua Dimensi Data <i>Space</i> (Kiri) ke Tiga Dimensi <i>Feature Space</i> (Kanan).....	24
Gambar 2.14	Ilustrasi Perhitungan C-Indeks .....	33
Gambar 3.1	Kerangka Konsep .....	45
Gambar 3.2	Diagram Alir Langkah Analisis .....	53
Gambar 4.1	Status Meninggal Pasien Kanker Serviks .....	62
Gambar 4.2	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Komplikasi .....	63
Gambar 4.3	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Anemia .....	64
Gambar 4.4	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Jenis Pengobatan .....	64
Gambar 4.5	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Stadium .....	65
Gambar 4.6	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Riwayat KB .....	65

Gambar 4.7	Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Pendidikan .....	66
Gambar 4.8	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Usia ....	67
Gambar 4.9	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Komplikasi .....	68
Gambar 4.10	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Status Anemia .....	69
Gambar 4.11	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Jenis Pengobatan .....	70
Gambar 4.12	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Stadium .....	71
Gambar 4.13	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Usia saat Menikah .....	72
Gambar 4.14	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Usia Pertama Kali Menstruasi .....	74
Gambar 4.15	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Siklus Menstruasi .....	75
Gambar 4.16	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Lama Menstruasi .....	76
Gambar 4.17	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Paritas .	77
Gambar 4.18	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Riwayat KB .....	78
Gambar 4.19	Kurva <i>Survival</i> Kaplan Meier Pasien Kanker Serviks Pendidikan .....	79
Gambar 4.20	Bazeline Hazard dari Model Cox PHM (X <sub>4</sub> dan X <sub>5</sub> ) .....	83
Gambar 4.21	Perbandingan Metode Cox PHM dan A-SURLSSVM pada Data Kanker Serviks .....	85
Gambar 4.22	Performansi pada Data Linier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100 dan 1000 .....	88
Gambar 4.23	Performansi pada Data Nonlinier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100 dan 1000 .....	90

Gambar 4.24	Performansi pada Data Linier dengan Jumlah Variabel Prediktor 6 dan 8 Jumlah Data 100 .....	91
Gambar 4.25	Performansi pada Data Nonlinier dengan Jumlah Variabel Prediktor 6 dan 8 Jumlah Data 100 .....	92
Gambar 4.26	Performansi pada Data Linier dan Nonlinier dengan Jumlah Variabel Prediktor 17 Jumlah Data 1000 .....	94
Gambar 4.27	Performansi pada Data Linier dan Nonlinier dengan Jumlah Variabel Prediktor Berbeda Jumlah Data 1000 .....	95
Gambar 4.28	Prosentase Setiap Variabel yang Tetap Ada dalam Model dengan 100 kali Replikasi .....	96
Gambar 4.29	Prosentase Kenaikan C-index setelah <i>Features Selection</i> dari 100 kali Replikasi .....	97



*(halaman ini sengaja dikosongkan)*

## DAFTAR TABEL

Tabel 2.1 Ilustrasi Perhitungan C-Indeks.....	32
Tabel 2.2 FIGO ( <i>The International Federation of Gynecology and Obstetrics</i> ) Stadium untuk Kanker Serviks .....	37
Tabel 2.3 FIGO ( <i>The International Federation of Gynecology and Obstetrics</i> ) Stadium untuk Kanker Serviks (Lanjutan).....	38
Tabel 3.1 Variabel Data Simulasi .....	46
Tabel 3.2 Skenario untuk Analisis Data Simulasi.....	47
Tabel 3.3 Skenario untuk Analisis Data Simulasi (Lanjutan).....	48
Tabel 3.4 Variabel Data Riil .....	49
Tabel 3.5 Variabel Data Riil (Lanjutan) .....	50
Tabel 3.6 Struktur Data Analisis <i>Survival</i> dengan <i>Time Independent Covariate</i> .....	51
Tabel 4.1 Statistika Deskriptif dari <i>Survival Time</i> .....	62
Tabel 4.2 Deskriptif Faktor $X_1, X_6, X_7, X_8, X_9, X_{10}$ .....	63
Tabel 4.3 Uji Logrank Faktor Usia .....	67
Tabel 4.4 Uji Logrank Faktor Komplikasi .....	68
Tabel 4.5 Uji Logrank Faktor Status Anemia .....	70
Tabel 4.6 Uji Logrank Faktor Jenis Pengobatan .....	71
Tabel 4.7 Uji Logrank Faktor Stadium .....	72
Tabel 4.8 Uji Logrank Faktor Usia Saat Menikah .....	73
Tabel 4.9 Uji Logrank Faktor Usia Pertama Kali Menstruasi .....	74
Tabel 4.10 Uji Logrank Faktor Siklus Menstruasi .....	75
Tabel 4.11 Uji Logrank Faktor Lama Menstruasi .....	76
Tabel 4.12 Uji Logrank Faktor Paritas .....	78
Tabel 4.13 Uji Logrank Faktor Riwayat KB .....	79
Tabel 4.14 Uji Logrank Faktor Pendidikan .....	80
Tabel 4.15 Cox PHM pada Data Kanker Serviks .....	81
Tabel 4.16 Asumsi Proportional Hazard Cox PHM pada Data Kanker Serviks ..	81
Tabel 4.17 Cox PHM pada Data Kanker Serviks .....	82
Tabel 4.18 Performansi Cox PHM pada Data Kanker Serviks .....	83
Tabel 4.19 Statistika Deskriptif dari Prognostik Indeks .....	84

Tabel 4.20 Performansi A-SURLSSVM pada Data Kanker Serviks .....	84
Tabel 4.21 Performansi A-SURLSSVM Hasil Seleksi Variabel .....	86
Tabel 4.22 Performansi A-SURLSSVM Hasil Seleksi Variabel Setelah Menghilangkan $X_2$ .....	87
Tabel 4.23 Performansi A-SURLSSVM Hasil Seleksi Variabel Setelah Menghilangkan $X_2$ dan $X_{11}$ .....	87
Tabel 4.24 Nilai C-index dari 100 Replikasi .....	98

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b>	Surat Keterangan Kelayakan Etik .....	107
<b>Lampiran 2.</b>	Syntax Membangkitkan Data Simulasi untuk $p=6$ pada Data dengan Pola Linier .....	108
<b>Lampiran 3.</b>	Syntax Membangkitkan Data Simulasi untuk $p=8$ pada Data dengan Pola Linier .....	109
<b>Lampiran 4.</b>	Syntax Membangkitkan Data Simulasi untuk $p=15$ pada Data dengan Pola Linier .....	110
<b>Lampiran 5.</b>	Syntax Membangkitkan Data Simulasi untuk $p=17$ pada Data dengan Pola Linier .....	112
<b>Lampiran 6.</b>	Syntax Membangkitkan Data Simulasi untuk $p=6$ pada Data dengan Pola Nonlinier .....	114
<b>Lampiran 7.</b>	Syntax Membangkitkan Data Simulasi untuk $p=8$ pada Data dengan Pola Nonlinier .....	115
<b>Lampiran 8.</b>	Syntax Membangkitkan Data Simulasi untuk $p=15$ pada Data dengan Pola Nonlinier .....	117
<b>Lampiran 9.</b>	Syntax Membangkitkan Data Simulasi untuk $p=17$ pada Data dengan Pola Nonlinier .....	119
<b>Lampiran 10.</b>	Data Pasien Kanker Serviks Rumah Sakit dr. Soetomo 2013-2016 .....	120
<b>Lampiran 11.</b>	Syntax untuk Menghitung Matriks D .....	123
<b>Lampiran 12.</b>	Syntax untuk Menghitung Nilai Fungsi Kernel .....	124
<b>Lampiran 13.</b>	Syntax Cox PHM dan Ukuran Performansinya .....	125
<b>Lampiran 14.</b>	Syntax untuk A-SURLSSVM dan Ukuran Performansinya .	127
<b>Lampiran 15.</b>	Syntax untuk Membuat Plot antara Performansi terhadap Persentase Sensoring .....	130
<b>Lampiran 16.</b>	Performansi untuk Data Simulasi $p=6$ dan $n=100$ pada Data dengan Pola Linier .....	131
<b>Lampiran 17.</b>	Performansi untuk Data Simulasi $p=6$ $n=100$ pada Data dengan Pola Nonlinier .....	132

<b>Lampiran 18.</b> Performansi untuk Data Simulasi $p=6$ $n=1000$ pada Data dengan Pola Linier .....	133
<b>Lampiran 19.</b> Performansi untuk Data Simulasi $p=6$ $n=1000$ pada Data dengan Pola Nonlinier .....	134
<b>Lampiran 20.</b> Performansi untuk Data Simulasi $p=8$ $n=100$ pada Data dengan Pola Linier .....	135
<b>Lampiran 21.</b> Performansi untuk Data Simulasi $p=8$ $n=100$ pada Data dengan Pola Nonlinier .....	136
<b>Lampiran 22.</b> Performansi untuk Data Simulasi $p=8$ $n=1000$ pada Data dengan Pola Linier .....	137
<b>Lampiran 23.</b> Performansi untuk Data Simulasi $p=8$ $n=1000$ pada Data dengan Pola Nonlinier .....	138
<b>Lampiran 24.</b> Performansi untuk Data Simulasi $p=15$ $n=100$ pada Data dengan Pola Linier .....	139
<b>Lampiran 25.</b> Performansi untuk Data Simulasi $p=15$ $n=100$ pada Data dengan Pola Nonlinier .....	140
<b>Lampiran 26.</b> Performansi untuk Data Simulasi $p=15$ $n=1000$ pada Data dengan Pola Linier .....	141
<b>Lampiran 27.</b> Performansi untuk Data Simulasi $p=15$ $n=1000$ pada Data dengan Pola Nonlinier .....	142
<b>Lampiran 28.</b> Performansi untuk Data Simulasi $p=17$ $n=100$ pada Data dengan Pola Linier .....	143
<b>Lampiran 29.</b> Performansi untuk Data Simulasi $p=17$ $n=100$ pada Data dengan Pola Nonlinier .....	144
<b>Lampiran 30.</b> Performansi untuk Data Simulasi $p=17$ $n=1000$ pada Data dengan Pola Linier .....	145
<b>Lampiran 31.</b> Performansi untuk Data Simulasi $p=17$ $n=1000$ pada Data dengan Pola Nonlinier .....	146
<b>Lampiran 32.</b> Performansi untuk Data Simulasi $p=6$ $n=5000$ pada Data dengan Pola Linier .....	147
<b>Lampiran 33.</b> Performansi untuk Data Simulasi $p=6$ $n=5000$ pada Data dengan Pola Nonlinier .....	148

<b>Lampiran 34.</b> Performansi untuk Data Simulasi $p=8$ $n=5000$ pada Data dengan Pola Linier .....	149
<b>Lampiran 35.</b> Performansi untuk Data Simulasi $p=8$ $n=5000$ pada Data dengan Pola Nonlinier .....	150
<b>Lampiran 36.</b> Performansi untuk Data Simulasi $p=15$ $n=5000$ pada Data dengan Pola Linier .....	151
<b>Lampiran 37.</b> Performansi untuk Data Simulasi $p=15$ $n=5000$ pada Data dengan Pola Nonlinier .....	152
<b>Lampiran 38.</b> Performansi untuk Data Simulasi $p=17$ $n=5000$ pada Data dengan Pola Linier .....	153
<b>Lampiran 39.</b> Performansi untuk Data Simulasi $p=17$ $n=5000$ pada Data dengan Pola Nonlinier .....	154
<b>Lampiran 40.</b> Performansi untuk Data Simulasi $n=100$ pada Data dengan Pola Linier .....	155
<b>Lampiran 41.</b> Performansi untuk Data Simulasi $n=1000$ pada Data dengan Pola Linier .....	156
<b>Lampiran 42.</b> Performansi untuk Data Simulasi $n=5000$ pada Data dengan Pola Linier .....	157
<b>Lampiran 43.</b> Performansi untuk Data Simulasi $n=100$ pada Data dengan Pola Nonlinier .....	158
<b>Lampiran 44.</b> Performansi untuk Data Simulasi $n=1000$ pada Data dengan Pola Nonlinier .....	159
<b>Lampiran 45.</b> Performansi untuk Data Simulasi $n=5000$ pada Data dengan Pola Nonlinier .....	160
<b>Lampiran 46.</b> Optimasi Parameter ASURLSSVM dan RBF kernel untuk Data Kanker Serviks .....	161
<b>Lampiran 47.</b> Syntax ASURLSSVM .....	165
<b>Lampiran 48.</b> Output A-SURLSSVM pada Data Pasien Kanker Serviks .....	166
<b>Lampiran 49.</b> Prognostik indeks pada Data Pasien Kanker Serviks .....	168

*(halaman ini sengaja dikosongkan)*

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis *survival* adalah suatu metode yang digunakan untuk menganalisis waktu suatu individu atau subjek mulai dari awal pengamatan sampai terjadinya kejadian. Kejadian yang diamati seperti kematian, timbul atau kambuhnya penyakit, rusak atau kembali bekerjanya sebuah mesin, dan kejadian pada individu yang berhenti atau mendapatkan pekerjaan (Kleinbaum & Klein, 2005).

Analisis *survival* banyak diterapkan pada bidang kesehatan, terutama pada penyakit kanker. Analisis *survival* juga mempertimbangkan penyensoran data. Data dikatakan tersensor apabila kejadian suatu individu tidak dapat diamati secara lengkap. Hal ini disebabkan oleh kejadian tak terduga yang mengakibatkan obyek keluar dari penelitian (Collet, 2003). Contohnya, subjek pengamatan tidak mengalami kejadian ketika penelitian berakhir dan subjek pengamatan menghilang selama penelitian. Jika pengamatan dilakukan sampai waktu terjadinya kejadian, misalnya kematian maka dikatakan data tidak tersensor.

Awalnya, model *survival* merupakan model parametrik yang didasari asumsi distribusi tertentu. Contohnya adalah model weibull, dengan model eksponensial sebagai kasus khusus (Van Belle *et al.*, 2010). Terobosan terbesar dalam pemodelan *survival* terjadi pada tahun 1972 ketika Cox mengusulkan modelnya yaitu Cox proporsional hazard model (Cox PHM) (Cox, 1972). Model Cox PHM adalah model *survival* yang semi-parametrik dan paling umum digunakan. Model Cox PHM diusulkan sebagai model *hazard* yaitu risiko mengalami kejadian pada waktu ke- $t$  dan kejadian itu tidak terjadi sebelumnya (Van Belle *et al.*, 2010). Karakter semi-parametrik model ini berasal dari *baseline hazard* yang tidak ditentukan dan tidak harus didasari asumsi distribusi tertentu. Kelebihan dari model ini adalah penggunaan fungsi Likelihood parsial yang terbukti sesuai dengan fungsi Likelihood.

Tujuan analisis *survival* adalah mengestimasi atau menginterpretasi fungsi *hazard* dan fungsi *survival* dari data *survival*. Kegunaan dari interpretasi fungsi



*hazard* adalah untuk mengetahui kelajuan dari suatu individu mengalami kejadian dalam interval  $t$  sampai  $t + \Delta t$ . Semakin besar kelajuan maka individu tersebut semakin berisiko mengalami kejadian. Tujuan kedua adalah membandingkan fungsi *survival* dan fungsi *hazard*, dan tujuan ketiga adalah untuk mengetahui hubungan antara waktu *survival* dengan variabel-variabel yang diduga mempengaruhi waktu *survival*. Hubungan tersebut dapat dimodelkan salah satunya dengan model regresi Cox PHM. Regresi Cox PHM mempunyai variabel terikat yaitu waktu *survival* dan variabel bebas yang merupakan variabel yang diduga berpengaruh terhadap waktu *survival*. Meskipun model *proportional hazard* telah umum dan banyak digunakan, namun model tersebut masih memiliki kelemahan. Kelemahan dari model ini adalah adanya asumsi proporsional (*hazard* berbeda namun proporsional antar kelompok pengamatan) dan asumsi linearitas dalam kovariat (Van Belle *et al.*, 2010).

Selama beberapa dekade terakhir, telah banyak metode yang diusulkan untuk memperbaiki kekurangan dari model *proportional hazard* yaitu kovariat yang nonlinear dan atau tanpa asumsi proporsional. Hastie dan Tibshirani (1990) memperkenalkan model aditif umum di mana bentuk linear diganti dengan *a sum of smooth functions*. Selain itu, analisis *survival* dengan pendekatan splines telah dilakukan penelitian oleh Sullivan (1988); Harrell *et al.* (1988); Gray (1992); dan Li & Luan (2005). Analisis *survival* dengan regresi kernel *Cox Proportional Hazard* juga telah diusulkan oleh Li & Luan (2003). Neural network untuk menerapkan model *Cox Proportional Hazard* pada efek kovariat non-linear dilakukan oleh Faraggi & Simon (1995); Bakker *et al.* (2004) dan Biganzoli *et al.* (1998) juga telah dilakukan. Biganzoli *et al.* (1998) mengenalkan model *Partial Logistic Neural Networks* (PLANN). Namun model PLANN yang diusulkan tersebut memiliki kelemahan yaitu datapoints perlu direplikasi di semua interval waktu. Penelitian dengan menggunakan pendekatan nonparametrik lainnya juga telah dilakukan oleh Inayati dan Purnami (2015) yaitu analisis *survival* nonparametrik pada pasien kanker serviks di RSUD dr. Soetomo Surabaya menggunakan metode kaplan meier dan uji Logrank. Selain itu, analisis *survival* dengan *Multivariate Adaptive Regression Spline* (MARS) oleh Respita (2017) juga telah dilakukan.

Namun pada penelitian saat ini umumnya data memiliki dimensi yang tinggi, seperti misalnya dalam kasus genetika komputasi. *Model Cox Proportional Hazard* serta model PLANN menghasilkan estimasi koefisien yang kurang akurat untuk data berdimensi tinggi. Dalam penelitian Suykens, Jak *et al.* (2002) diketahui bahwa metode berbasis kernel SVM dan LS-SVM sangat cocok untuk menangani data dimensi tinggi. LS-SVM lebih baik dibandingkan dengan SVM standart dalam hal proses perhitungan, konvergensi cepat, dan presisi yang tinggi. Jika SVM menggunakan fungsi *constrain* yang hanya pertidaksamaan maka LS-SVM diformulasikan menggunakan fungsi *constrain* berupa persamaan. Sehingga solusi LS-SVM dihasilkan dengan menyelesaikan persamaan linier. Hal ini berbeda dengan SVM, dimana solusinya dihasilkan melalui penyelesaian *quadratic programming* (Suykens dan Vandewalle, 1999a, 1999b, 1999c; Bhavsar, *et al.*, 2012; Trapsilasiwi, 2013).

Van Belle *et al.* (2007) dan (2008) menerapkan analisis *survival* yang memiliki kovariat yang non-linear dengan pendekatan *Support Vector Machine* (SVM). Van Belle *et al.* (2008) menggunakan *survival* SVM untuk membandingkan performansi metode dengan menggunakan Cox PHM, *Accelerated Failure Time Model* (AFT model), cSVM (SVM dengan linear kernel), dan cSVM- Gaussian Radial Basis (SVM dengan RBF kernel). Dengan menggunakan dua data set, didapatkan kesimpulan bahwa hasil dengan menggunakan SVM lebih baik dari pada menggunakan Cox PHM maupun AFT. Van Belle *et al.* (2010) meneliti tentang analisis *survival* untuk pasien kanker payudara dengan membandingkan beberapa metode. Diantaranya dengan Cox PHM, *Partial Logistic Artificial Neural Networks* dengan *Automatic Relevance Determination* (PLANN-ARD), *Survival SVM* (SURSVM), *Survival Least Square SVM* (SURLSSVM) dan *Additive Survival Least Square SVM* (A-SURLSSVM). Pada penelitian tersebut, didapatkan kesimpulan bahwa metode A-SURLSSVM merupakan model terbaik karena memiliki ukuran variasi ( $R^2$ ) tertinggi dan nilai kerugian (*loss*) paling kecil.

Kanker serviks (kanker leher rahim) merupakan kanker yang paling umum kedua di dunia pada perempuan, dengan perkiraan 471.000 kasus baru (233.000 meninggal dunia) pada tahun 2000. Dari data tersebut, lebih dari 80% penderita

berasal dari negara berkembang di Asia Selatan, Asia Tenggara, Sub-Sahara Afrika, Amerika Tengah, dan Amerika Selatan (WHO, 2005). Di Indonesia kanker serviks menempati peringkat kedua setelah kanker payudara dari segi jumlah penderita kanker pada perempuan. Namun, sebagai penyebab kematian masih menempati peringkat pertama (Ocviyanti & Handoko, 2013). Ketidaktahuan para perempuan akan ancaman kanker serviks turut membantu banyaknya perempuan yang meninggal akibat penyakit ini (Murniati & Lisuwarni, 2014). Berdasarkan data WHO pada tahun 2008, terdapat 530.202 kasus baru kanker serviks di seluruh dunia. Dengan jumlah itu berarti diperkirakan akan didapatkan 1 kasus baru kanker serviks setiap menitnya di dunia. Secara keseluruhan diperkirakan insidensi kanker serviks di seluruh dunia adalah sebesar 16,2 per 100.000 penduduk (Ocviyanti & Handoko, 2013).

Penelitian analisis *survival* pada kanker serviks telah banyak dilakukan. Wijayanti (2014) melakukan penelitian perbandingan analisis Cox PHM dan analisis *survival* bayesian pada data penderita kanker serviks di RSUD dr. Soetomo dimana pada penelitian ini menyimpulkan bahwa variabel prediktor yang berpengaruh terhadap ketahanan penderita kanker serviks di RSUD dr. Soetomo adalah stadium dan jenis kanker. Novitasari (2014) melakukan penelitian analisis *survival* pada data rekurensi dengan menggunakan *counting process approach* dan model PWP-GT pada data kanker serviks dimana pada penelitian ini menyimpulkan bahwa variabel prediktor yang berpengaruh rekurensi pada penderita kanker serviks di RSUD dr. Soetomo Surabaya adalah stadium dan jenis kanker. Wulansari (2015) juga melakukan penelitian analisis *survival* pada data rekurensi dengan pendekatan Andersen/Gill pada data pasien kanker serviks di RSUD dr. Soetomo Surabaya. Pada penelitian tersebut didapatkan kesimpulan bahwa variabel stadium, komplikasi, dan berat badan secara parsial berpengaruh untuk mengalami *event* atau rekurensi. Penelitian analisis *survival* pada kanker serviks telah dilakukan dengan pendekatan parametrik dan semiparametrik. Berdasarkan penjelasan sebelumnya, bahwa pendekatan parametrik dan semiparametrik memiliki beberapa kelemahan yaitu adanya asumsi-asumsi yang harus dipenuhi.

Peningkatan performansi dari suatu model juga dapat dilakukan dengan *features selection*. *Features selection* adalah suatu kegiatan yang umumnya bisa

dilakukan secara preprocessing dan bertujuan untuk memilih *feature* yang berpengaruh dan mengesampingkan *feature* yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data (Chandrashekar dan Sahin, 2014).

Pada penelitian ini akan digunakan analisis *survival* dengan pendekatan nonparametrik yaitu analisis *survival* dengan metode *Additive Survival Least Square SVM* (A-SURLSSVM) pada berbagai jumlah data baik untuk data dengan pola linier maupun nonlinier. Performansi dan hasil prediksi A-SURLSSVM akan dibandingkan dengan metode analisis *survival* yang lain yaitu Cox PHM dengan menggunakan *hazard ratio*, Logrank, dan C-index. Ketiga performansi tersebut dihitung dari prognostik indeks masing-masing metode. Prognostik indeks merupakan pengganti dari fungsi *hazard*, dimana kesesuaian antara observasi waktu kegagalan dan *output* model dioptimalkan. *Output* model merupakan fungsi prognostik juga disebut sebagai fungsi utilitas dan lebih spesifik dalam penelitian medis disebut prognostik indeks atau fungsi kesehatan. Data yang digunakan dalam penelitian ini merupakan data simulasi dan data riil. Data riil yang digunakan merupakan data pasien kanker serviks di rumah sakit dr. Soetomo Surabaya. Sedangkan data simulasi digunakan untuk mengetahui performansi metode yang digunakan dalam berbagai skenario. Selain itu, pada penelitian ini juga menggunakan *features selection* dengan pendekatan metode *backward elimination* untuk mengetahui variabel-variabel yang dapat meningkatkan performansi.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah diuraikan, maka dapat dirumuskan permasalahan pada penelitian ini adalah bagaimana mengkaji metode A-SURLSSVM untuk mendapatkan persamaan prognostik indeks, mengaplikasikan metode tersebut pada data simulasi dan studi kasus pasien kanker serviks di RSUD dr. Soetomo Surabaya serta membandingkan performansi metode A-SURLSSVM dengan metode Cox PHM dengan dan tanpa *features selection*.

## **1.3 Tujuan**

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengkaji metode A-SURLSSVM untuk mendapatkan prognostik indeks.
2. Mengaplikasikan metode A-SURLSSVM pada data simulasi dan studi kasus pasien kanker serviks di RSUD dr. Soetomo Surabaya.
3. Mengetahui perbandingan performansi metode A-SURLSSVM dengan metode Cox PHM dengan dan tanpa *features selection*.

#### **1.4 Manfaat**

Melalui penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Menambah wawasan mengenai analisis *survival* untuk data kanker serviks maupun data simulasi menggunakan *additive Survival Least Squarel SVM* (A-SURLSSVM).
2. Memberikan informasi tentang prediksi pasien kanker serviks yang masuk kategori beresiko tinggi atau rendah untuk terjadi *event* (meninggal) berdasarkan prognostik indeks.

#### **1.5 Batasan Masalah**

Batasan masalah yang digunakan agar penelitian dapat berjalan secara terarah serta dapat mencapai tujuan yang diinginkan adalah sebagai berikut:

1. Sensor yang digunakan dalam penelitian ini merupakan sensor kanan (*right censor*) yang berarti bahwa jika pasien kanker serviks belum mengalami *event* sampai dengan masa pendataan selesai. Pada penelitian ini digunakan data pasien kanker serviks yang pernah menjalani pengobatan di RSUD dr. Soetomo Surabaya Januari tahun 2013 sampai dengan Desember tahun 2016.
2. Distribusi *survival time* antara data simulasi dan data riil tidak disamakan karena pengambilan data riil memerlukan waktu yang lama, sedangkan studi simulasi harus segera dilakukan.
3. Fungsi kernel yang digunakan adalah fungsi kernel Gaussian Radial Basis (RBF).

## **BAB 2**

### **TINJAUAN PUSTAKA**

Tinjauan pada bab ini dibagi menjadi dua bagian, yaitu tinjauan statistika dan tinjauan non statistika. Tinjauan statistika membahas tentang analisis *survival*, Kurva *Survival* Kaplan Meier dan uji Logrank, *Cox proportional hazard*, *Support Vector Machine* (SVM), *Least Square Vector Machine* (LS-SVM), *Survival SVM* (SUR SVM), *Additive Survival Least Square SVM* (A-SURLSSVM), kriteria kebaikan model, pembangkitan data *survival time*, *features selection* dengan *backward elimination*. Sedangkan tinjauan non statistik membahas tentang kanker serviks.

#### **2.1 Analisis Survival**

Analisis *survival* adalah suatu metode untuk menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* atau *start point* sampai dengan terjadinya suatu kejadian khusus atau *end point* (Collett, 2003). Tujuan dari analisis *survival* adalah menaksir probabilitas kelangsungan hidup, kesembuhan, kematian, kekambuhan dan peristiwa lainnya dalam periode waktu tertentu. Selain itu, analisis *survival* juga dapat digunakan untuk mengetahui hubungan variabel dependen dan independen, dengan variabel dependen berupa waktu *survival*. Waktu *survival* diperoleh dari suatu pengamatan terhadap objek yang dicatat waktu dari awal kejadian sampai terjadinya peristiwa tertentu, yaitu kegagalan dari setiap objek yang disebut dengan *failure event* (Collett, 2003).

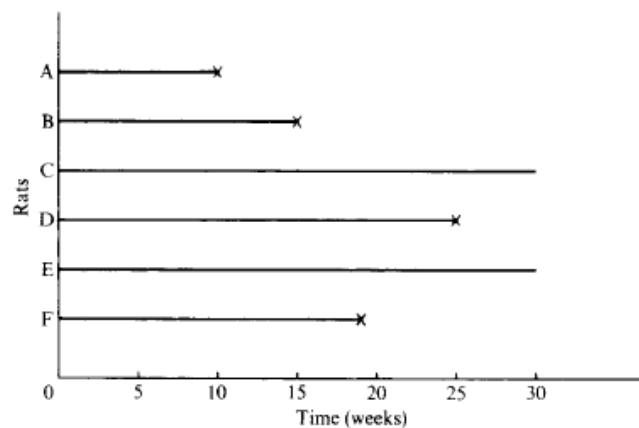
Dalam menentukan waktu *survival*, terdapat tiga elemen yang harus diperhatikan, yaitu : waktu awal (*time origin*), definisi *failure time* keseluruhan yang harus jelas dan skala waktu (*measurement scale of time*) sebagai satuan pengukuran yang harus jelas (Le, 1997). Analisis *survival* memiliki ciri khusus, yaitu distribusi data lama waktu (*lifetime*) yang bersifat menceng (*skew*) kanan karena nilainya akan selalu positif dan karena adanya data tersensor (Lee & Wang, 2003). Secara umum, ada tiga alasan penyebab terjadinya penyensoran menurut Collet (2003) adalah sebagai berikut:

1. *Loss to follow up*, yaitu sensor yang terjadi bila objek pindah, hilang dari pengamatan, atau menolak untuk berpartisipasi.
2. *Drop out*, yaitu sensor yang terjadi bila perlakuan dihentikan karena suatu alasan tertentu.
3. *Termination of study*, yaitu sensor yang terjadi bila masa penelitian berakhir sementara objek yang diobservasi belum mencapai *failure event*.

Terdapat beberapa cara pengklasifikasian tipe data tersensor (Lee & Wang, 2003) yaitu :

1. Tersensor tipe I

Data tersensor tipe I terjadi jika semua objek penelitian masuk ke dalam penelitian pada waktu yang sama dan periode penelitian telah ditentukan oleh peneliti.

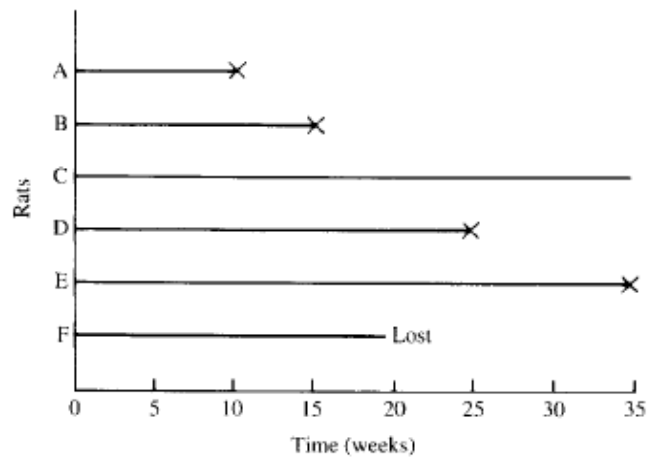


**Gambar 2.1** Contoh Data Tersensor Tipe I (Lee & Wang, 2003)

Misalkan kasus yang terjadi pada Gambar 2.1, dalam kurun waktu 30 minggu akan diamati mengenai perkembangan ukuran tumor. Ternyata, selama periode 30 minggu tersebut, terjadi perkembangan ukuran tumor pada *rats* A, B dan D dalam masa pengamatan. Untuk *rats* C dan E tidak terjadi perkembangan ukuran tumor hingga waktu pengamatan berakhir (minggu ke-30), sedangkan untuk *rats* F meninggal dunia sebelum minggu ke-20 yang penyebabnya bukan perkembangan ukuran tumor. Data *rats* A, B, dan D yang mengalami perkembangan ukuran tumor dan meninggal disebut data tidak tersensor dan data lainnya disebut data tersensor.

## 2. Tersensor tipe II

Terjadi jika pengamatan berhenti setelah  $d$  pengamatan dengan banyaknya pengamatan yang harus mencapai *failure* yang ditentukan sejak awal. Pada data tersensor tipe II, individu masuk ke dalam penelitian pada waktu yang sama dan penelitian dihentikan jika sejumlah individu yang telah ditentukan gagal, mati, atau rusak.



**Gambar 2.2** Contoh Data Tersensor Tipe II (Lee & Wang, 2003)

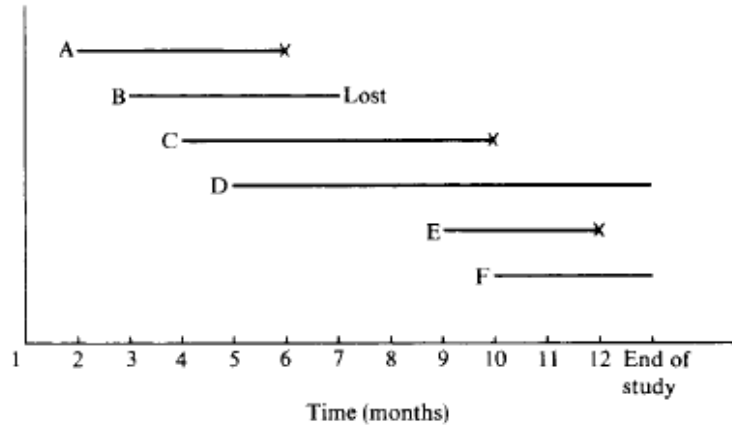
Misalkan kasus yang terjadi pada Gambar 2.2, dalam kurun waktu 35 minggu akan diamati mengenai perkembangan ukuran tumor dan penelitian baru akan dihentikan jika dua *rats* meninggal dunia. Ternyata, sampai pada minggu ke-35, terdapat dua *rats* meninggal dunia (C dan E) sehingga penelitian dihentikan pada minggu ke-35. Seperti terlihat pada Gambar bahwa *rat F lost to follow up*.

## 3. Tersensor tipe III

Data tersensor tipe III disebut juga tersensor secara progresif terjadi jika setiap individu masuk ke dalam penelitian pada waktu yang berbeda-beda selama periode penelitian. Data ini seringkali terdapat pada penelitian klinis. Misalkan kasus yang terjadi pada Gambar 2.3, dimana selama periode penelitian terdapat enam pasien yang masuk ke dalam pengamatan. Pasien A masuk ke dalam pengamatan pada bulan ke-2, pasien B masuk pada bulan ke-3, pasien C masuk pada bulan ke-4, dan seterusnya. Pada akhir periode penelitian, diketahui bahwa 3 pasien meninggal dunia (A, C, dan E), pasien



B *lost to follow up* dan pasien D dan F *withdrawn alive* hingga akhir pengamatan.

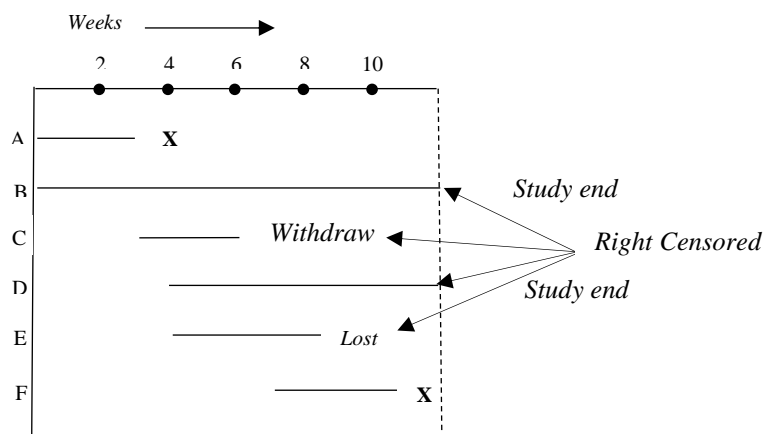


**Gambar 2.3** Contoh Data Tensensor Tipe III (Lee & Wang, 2003)

Jenis pengklasifikasian lain untuk data tensensor (Collett, 2003) adalah :

1. Sensor kanan

Dikatakan tensensor kanan jika objek yang diamati pindah atau keluar sebelum *failure* terjadi dan setelah data dikumpulkan atau penelitian selesai, objek belum mengalami *failure*. Dengan kata lain, data tensensor kanan terjadi bila waktu *survival* seorang individu tidak lengkap sampai dengan periode *follow-up*. Berikut ini adalah ilustrasi dari data tensensor kanan.

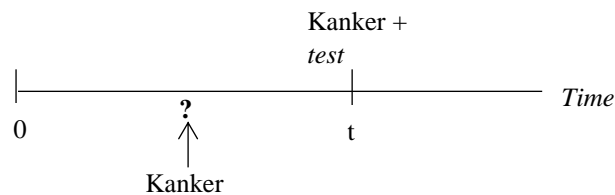


**Gambar 2.4** Ilustrasi Data Sensor Kanan (Kleinbaum & Klein, 2012)

2. Sensor kiri

Dikatakan tensensor kiri jika terjadi *failure* sebelum data mulai dikumpulkan. Dengan kata lain, data tensensor kiri terjadi bila waktu *survival* seorang individu kurang atau sama dengan waktu *survival* individu

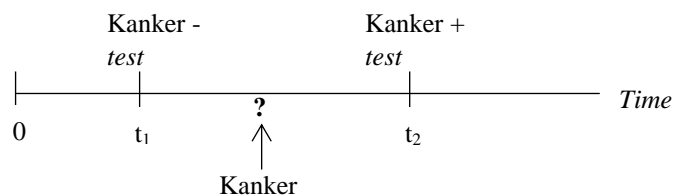
saat diobservasi atau dapat dikatakan bahwa saat dilakukan observasi, individu sudah mengalami *failure* tetapi tidak diketahui secara pasti kapan *failure* itu terjadi. Sebagai contoh, peneliti mengamati pasien penyakit kanker, peneliti dapat mencatat kejadian tepatnya seseorang tersebut positif kanker di tes pertamanya, namun peneliti tidak memiliki catatan tentang waktu tepatnya seseorang tersebut mulai berpenyakit kanker, dengan demikian pasien kanker tersebut tersensor kiri yaitu ketika mengalami kejadian pertama dengan hasil positif kanker. Berikut ini merupakan ilustrasi dari data sensor kiri.



**Gambar 2.5** Ilustrasi Data Sensor Kiri (Kleinbaum & Klein, 2012)

3. Sensor interval

Dapat terjadi jika pemutusan dalam pengumpulan data dan objek penelitian mengalami *failure* diantara interval waktu tertentu. Sebagai contohnya, jika catatan medis menunjukkan bahwa pada usia 45 tahun pasien kanker dalam contoh di atas kondisinya sehat dan belum berpenyakit kanker, kemudian pasien melakukan tes pertama saat berumur 50 tahun dan terdiagnosis terkena penyakit kanker, dengan demikian usia saat didiagnosis positif kanker adalah antara 45 dan 50 tahun. Berikut ini adalah ilustrasi dari data sensor interval.



**Gambar 2.6** Ilustrasi Data Sensor Interval (Kleinbaum & Klein, 2012)

Menurut Collett (2003), pada analisis *survival* terdapat dua macam fungsi utama yaitu fungsi *survival* dan fungsi *hazard*. Apabila  $T$  merupakan variabel random yang melambangkan waktu *survival* dan mempunyai fungsi distribusi

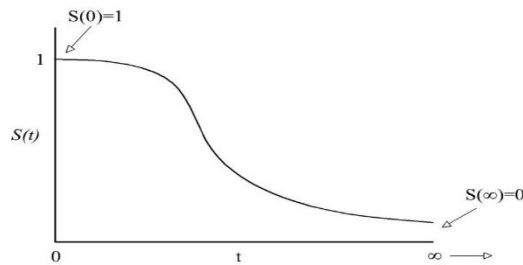
peluang  $f(t)$ , maka fungsi kepadatan peluang dan fungsi distribusi kumulatif dapat dinyatakan pada persamaan (2.1).

$$f(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t} \quad (2.1)$$

$$F(t) = P(T \leq t) = \int_0^t f(u) du \quad (2.2)$$

Fungsi di atas menyatakan bahwa probabilitas waktu *survival* lebih kecil dari suatu nilai  $t$ . Fungsi *survival*  $S(t)$ , didefinisikan sebagai probabilitas bahwa waktu *survival* lebih besar atau sama dengan  $t$ , sehingga

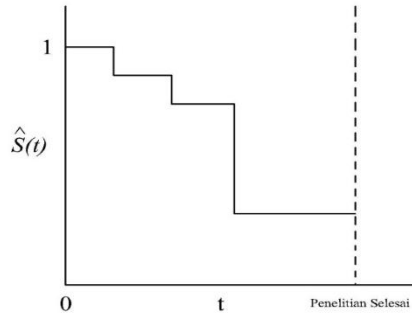
$$S(t) = P(T > t) = 1 - F(t) = 1 - P(T \leq t) \quad (2.3)$$



**Gambar 2.7** Ilustrasi  $S(t)$  secara Teori

Secara teori, saat  $t$  berada pada rentang 0 sampai tak terhingga, fungsi *survival* akan membentuk kurva seperti pada Gambar 2.7. Seperti pada Gambar 2.7 tersebut, saat  $t$  merupakan identitas dari sumbu x, maka semua fungsi *survival* mengikuti karakteristik sebagai berikut:

1. Tidak meningkat, pada saat nilai  $t$  meningkat maka kurva akan turun.
2. Pada waktu  $t = 0$ ,  $S(t) = S(0) = 1$ . Pada saat mulai penelitian, karena belum ada pasien yang dinyatakan meninggal maka peluang dari bertahan pada waktu 0 adalah satu.
3. Pada waktu  $t = \infty$ ,  $S(t) = S(\infty) = 0$ . Secara teori jika periode penelitian meningkat tanpa batas, pada akhirnya tidak ada pasien yang bertahan.



**Gambar 2.8** Ilustrasi  $S(t)$  pada Penerapan

Saat menggunakan data yang sebenarnya, sering terbentuk grafik seperti pada Gambar 2.8 di atas. Grafik tersebut terjadi karena periode penelitian terbatas maka memungkinkan pasien meninggal karena adanya penyebab lain (*competing risk*).

Fungsi *hazard* merupakan suatu laju kegagalan atau *failure* sesaat dengan asumsi bahwa suatu objek mencapai kejadian khusus sampai waktu ke- $t$ , dengan syarat telah bertahan sampai waktu tersebut. Misal probabilitas variabel random  $T$ , berada diantara  $t$  dan  $t + \Delta t$ , dengan syarat  $T$  lebih besar  $t$  dan dapat ditulis  $P(t \leq T < t + \delta | T > t)$  sehingga fungsi *hazard* yang diperoleh adalah

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T > t)}{\Delta t} \quad (2.4)$$

Berdasarkan definisi di atas, dapat diperoleh hubungan antara fungsi *survival* dan fungsi *hazard* dengan menggunakan teori probabilitas bersyarat  $P(A|B) = P(A \cap B)/P(B)$ , dimana  $P(A|B)$  adalah suatu probabilitas kejadian bersama antara A dan B. Nilai probabilitas bersyarat dari definisi fungsi *hazard* dinyatakan pada persamaan (2.5).

$$\frac{P(t \leq T < t + \Delta t)}{P(T \geq t)} = \frac{F(t + \Delta t) - F(t)}{S(t)} \quad (2.5)$$

Dengan  $F(t)$  adalah fungsi distribusi dari  $T$  sehingga,

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \quad (2.6)$$

Dengan  $F'(t) = f(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\}$  merupakan turunan dari  $F(t)$ .

Sehingga hubungan antara fungsi *survival* dengan dengan fungsi *hazard* dinyatakan pada persamaan (2.7).

$$h(t) = \frac{f(t)}{S(t)} \quad (2.7)$$

$$= \frac{d(F(t))}{dt}$$

$$= \frac{d(1 - S(t))}{d(t)}$$

$$= \frac{-d(S(t))}{d(t)}$$

$$h(t) d(t) = \frac{-d(S(t))}{S(t)}$$

$$\int_0^t h(t) d(t) = -\int_0^t \frac{1}{S(t)} d(S(t))$$

$$-\int_0^t h(t) d(t) = \int_0^t \frac{1}{S(t)} d(S(t))$$

$$-H(t) = \ln S(t)$$

$$S(t) = \exp\{-H(t)\}$$

Fungsi  $H(t)$  adalah *hazard* kumulatif yang diperoleh dari fungsi *survival*.

Hubungan di antara fungsi kumulatif *hazard* dan fungsi *survival* dinyatakan pada persamaan (2.8).

$$H(t) = -\ln S(t) \quad (2.8)$$

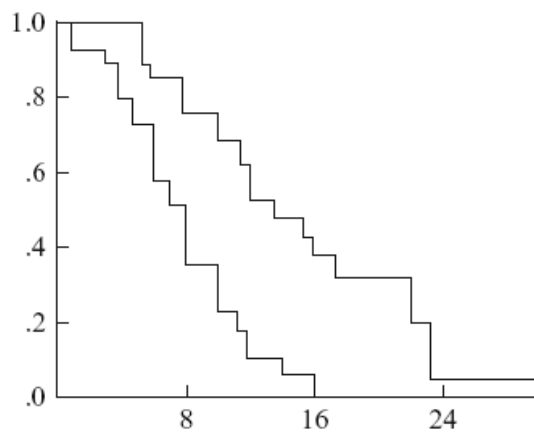
## 2.2 Kurva *Survival* Kaplan Meier dan Uji Logrank

Menurut Kleinbaum dan Klein (2012), analisis Kaplan Meier digunakan untuk menaksir fungsi *survival*. Apabila probabilitas dari Kaplan Meier adalah  $\hat{S}(t_{(j)})$  maka persamaan umumnya ditunjukkan pada persamaan (2.9).

$$\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \hat{P}(T > t_{(j)} | T \geq t_{(j)}) \quad (2.9)$$

$$\hat{S}(t_{(j-1)}) = \prod_{i=1}^{j-1} \hat{P}(T > t_{(i)} | T \geq t_{(i)}) \quad (2.10)$$

Berdasarkan hasil estimasi fungsi *survival* pada persamaan (2.9), dapat dibentuk kurva *survival* Kaplan Meier yaitu suatu kurva yang menggambarkan hubungan antara estimasi fungsi *survival* dengan waktu *survival*. Berikut merupakan ilustrasi dari kurva *survival* Kaplan Meier.



**Gambar 2.9** Ilustrasi Kurva *Survival* Kaplan Meier ( Kleinbaum, D. G., & Klein, M. 2012)

Uji Logrank merupakan uji yang digunakan untuk membandingkan kurva *survival* dalam grup yang berbeda (Kleinbaum & Klein, 2012).

Hipotesis dari uji Logrank untuk dua grup atau lebih adalah sebagai berikut:

$H_0$  : tidak ada perbedaan kurva *survival* dalam grup yang berbeda

$H_1$  : paling sedikit ada satu perbedaan kurva *survival* dalam grup yang berbeda

Statistik Uji :

$$\chi^2 = \sum_{h=1}^G \frac{(O_h - E_h)^2}{E_h} \quad (2.11)$$

Dengan

$$O_h - E_h = \sum_{j=1}^G (m_{hj} - e_{hj}) \text{ dan } e_{hj} = \left( \frac{n_{hj}}{\sum_{h=1}^G n_{hj}} \right) \left( \sum_{h=1}^G m_{hj} \right)$$

$m_{hj}$  = jumlah subjek yang gagal dalam grup ke- $h$  pada waktu  $t_{(j)}$

$n_{hj}$  = jumlah subjek yang beresiko gagal seketika pada grup ke- $h$  sebelum waktu  $t_{(j)}$

$e_{hj}$  = nilai ekspektasi dalam grup ke- $h$  pada waktu  $t_{(j)}$

$G$  = jumlah grup

Tolak  $H_0$  jika nilai  $\chi^2 > \chi_{\alpha, G-1}^2$

### 2.3 Regresi Cox *Proportional Hazard*

Regresi Cox *proportional hazard* atau dikenal dengan regresi Cox merupakan salah satu analisis *survival* yang sering digunakan. Metode ini pertama kali dikenalkan oleh Cox (1972). Regresi Cox *proportional hazard* ini tidak mempunyai asumsi mengenai sifat dan bentuk sesuai dengan distribusi normal seperti asumsi pada regresi yang lain, yang digunakan sesuai dengan respon yang digunakan.

#### 2.3.1 Model *Proportional Hazard*

Nilai variabel prediktor pada model *proportional hazard* dinyatakan oleh vektor  $\mathbf{x}$ , dengan  $\mathbf{x}=(x_1, x_2, \dots, x_p)^T$ . Fungsi *baseline hazard* dinyatakan sebagai  $h_0(t)$  merupakan fungsi *hazard* untuk tiap-tiap individu dimana semua variabel prediktor dalam vektor  $\mathbf{x}$  bernilai 0 (Collett, 2003).

Fungsi *hazard* untuk individu ke- $i$  dapat ditulis sebagai berikut:

$$h_i(t, x) = h_0(t)\psi(x_i) \tag{2.12}$$

Dengan  $\psi(x_i)$  adalah fungsi dari vektor variabel prediktor untuk individu ke- $i$ . Fungsi  $\psi(x_i)$  dapat diinterpretasikan sebagai fungsi resiko seseorang pada waktu ke- $t$  dengan vektor variabel prediktor  $x_i$  relatif terhadap risiko dari suatu

individu yang mempunyai  $x=0$ . Secara umum, oleh Le (1997), model *proportional hazard* diberikan dalam persamaan (2.13).

$$h(t, \mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}) \quad (2.13)$$

atau dapat dijabarkan dalam persamaan (2.14).

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) \quad (2.14)$$

### 2.3.2 Estimasi Parameter Regresi Cox PHM

Model terbaik didapatkan dengan melakukan estimasi koefisien variabel independen  $x_1, x_2, \dots, x_p$  yaitu  $\beta_1, \beta_2, \dots, \beta_p$ . Koefisien  $\beta$  pada model Cox PHM dapat diestimasi menggunakan metode *Maksimum Likelihood* (MLE). Apabila terdapat  $n$  sampel, dengan urutan waktu *failure*  $t_{(1)} < t_{(2)} < \dots < t_{(n)}$  dan  $R(t_i)$  merupakan himpunan waktu yang berisiko pada waktu  $t_{(i)}$  yang terdiri dari semua individu dengan waktu *survival*-nya paling sedikit. Fungsi parsial *likelihood* untuk model Cox PHM dinyatakan pada persamaan (2.15).

$$L(\beta) = \prod_{t=1}^n \frac{\exp(\boldsymbol{\beta}^T \mathbf{x}_t)}{\sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_l)} \quad (2.15)$$

Fungsi *ln-likelihood* dari persamaan (2.14) adalah sebagai berikut:

$$\begin{aligned} \ln L(\beta) &= \sum_{t=1}^n (\boldsymbol{\beta}^T \mathbf{x}_t) - \sum_{t=1}^n \ln \left[ \sum_{l \in R(t_i)} \exp(\boldsymbol{\beta}^T \mathbf{x}_l) \right] \\ &= \sum_{t=1}^n \sum_{i=1}^p (\beta_i x_{it}) - \sum_{t=1}^n \ln \left[ \sum_{l \in R(t_i)} \exp \sum_{i=1}^p (\beta_i x_{il}) \right] \\ &= \sum_{t=1}^n \left[ \sum_{i=1}^p (\beta_i x_{it}) - \ln \left[ \sum_{l \in R(t_i)} \exp \sum_{i=1}^p (\beta_i x_{il}) \right] \right] \end{aligned} \quad (2.16)$$

Berdasarkan persamaan (2.15), maka dapat dicari turunan pertama dan kedua.

$$\frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_1} = \sum_{t=1}^n \left[ x_{1t} - \frac{\sum_{l \in R(t_i)} x_{1l} \exp \sum_{i=1}^p (\beta_i x_{il})}{\sum_{l \in R(t_i)} \exp \sum_{i=1}^p (\beta_i x_{il})} \right] \quad (2.17)$$



$$\frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_2} = \sum_{t=1}^n \left[ x_{2t} - \frac{\sum_{i \in R(t)} x_{2t} \exp \sum_{i=1}^p (\beta_i x_{it})}{\sum_{i \in R(t)} \exp \sum_{i=1}^p (\beta_i x_{it})} \right]$$

$$\vdots$$

$$\frac{\partial \ln L(\beta_1, \beta_2, \dots, \beta_p)}{\partial \beta_p} = \sum_{t=1}^n \left[ x_{pt} - \frac{\sum_{i \in R(t)} x_{pt} \exp \sum_{i=1}^p (\beta_i x_{it})}{\sum_{i \in R(t)} \exp \sum_{i=1}^p (\beta_i x_{it})} \right]$$

$$\frac{\partial^2 L(\beta)}{\partial \beta_j \partial \beta_{j^*}} = - \sum_{t=1}^n \frac{\sum_{i \in R(t)} x_{jt} x_{j^*t} \exp \sum_{i=1}^p (\beta_i x_{it})}{\sum_{i \in R(t)} \exp \sum_{i=1}^p (\beta_i x_{it})} + \sum_{t=1}^n \left( \frac{\sum_{i \in R(t)} x_{jt} \exp \sum_{i=1}^p (\beta_i x_{it})}{\sum_{i \in R(t)} \exp \sum_{i=1}^p (\beta_i x_{it})} \frac{\sum_{i \in R(t)} x_{j^*t} \exp \sum_{i=1}^p (\beta_i x_{it})}{\sum_{i \in R(t)} \exp \sum_{i=1}^p (\beta_i x_{it})} \right)$$

dengan  $j, j^* = 1, 2, \dots, p$

Iterasi numerik dengan metode Newton-Raphson digunakan karena turunan pertama implisit (Collet, 2003). Jika  $\mathbf{g}(\beta)$  merupakan vektor berukuran  $p \times 1$  yang merupakan turunan pertama fungsi  $\ln L(\beta)$  terhadap parameter  $\beta$ .  $\mathbf{H}(\beta)$  merupakan matrik hessian berukuran  $p \times p$  yang berisi turunan kedua dari fungsi *ln-likelihood*, maka estimasi parameter pada iterasi ke  $(l+1)$  adalah  $\beta^{(l+1)} = \beta^{(l)} - \mathbf{H}^{-1}(\beta^{(l)}) \mathbf{g}(\beta^{(l)})$ . Iterasi akan berhenti jika  $\|\beta^{(l+1)} - \beta^{(l)}\| \leq \varepsilon$ , dengan  $\varepsilon$  merupakan suatu bilangan yang sangat kecil.

### 2.3.3 Pengujian Signifikansi Parameter

Pada analisis data *survival* ada dua pengujian parameter yang dilakukan, yaitu secara serentak dan parsial. Distribusi sta-tistik ujinya adalah *Chi Square*. Berikut ini pengujian parameter yang dilakukan setelah mendapatkan model (Le, 1997).

#### 1. Pengujian serentak

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

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, j = 1, 2, \dots, p$$

Statistik uji

$$\text{Uji rasio } \textit{likelihood}: G^2 = -2 \ln \Lambda$$

dengan:  $\Lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$

Tolak  $H_0$  jika  $G_{hit}^2 > \chi_{p,\alpha}^2$

$L(\hat{\omega})$  merupakan nilai *likelihood* untuk model tanpa menyertakan variabel prediktor dan  $L(\hat{\Omega})$  merupakan nilai *likelihood* untuk model lengkap dengan menyertakan semua variabel prediktor.

## 2. Pengujian parsial

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik uji

$$\text{Uji Wald : } W^2 = \frac{(\hat{\beta}_j)^2}{(\text{SE}(\hat{\beta}_j))^2} \quad (2.18)$$

$$\text{dengan, } \text{SE}(\hat{\beta}_j) = \sqrt{\text{var}(\hat{\beta}_j)}, \text{ var}(\hat{\beta}) = -H^{-1}(\hat{\beta}) \quad (2.19)$$

Tolak  $H_0$  jika  $W_{hit}^2 > \chi_{1,\alpha}^2$

$\text{var}(\hat{\beta}_j)$  didapatkan dari elemen diagonal ke  $j+1$  dari  $\text{var}(\hat{\beta})$ .

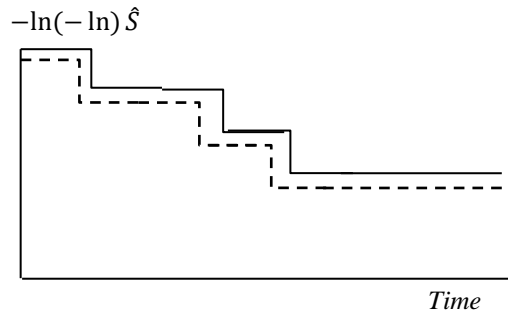
### 2.3.4 Asumsi *Proportional Hazard*

Hal yang menarik pada regresi Cox adalah data tidak harus memenuhi distribusi apapun. (Hosmer, Lemeshow & May, 2008). Sedangkan asumsi pemodelan yang harus dipenuhi dengan regresi Cox adalah bahwa fungsi *hazard* harus *proportional* setiap waktu karena regresi Cox tidak mengakomodasi variabel yang berubah-ubah sepanjang waktu (Collett, 2003). Asumsi *proportional* adalah variabel prediktor independen terhadap waktu dan hubungan antara kumulatif *hazard* konstan setiap waktu. Kleinbaum & Klein (2005) menyatakan ada tiga pendekatan umum untuk mengevaluasi asumsi *proportional hazard* yaitu.

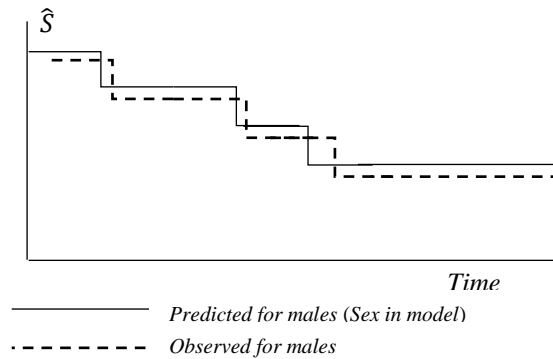
a. Grafik

Asumsi *proportional hazard* salah satunya dapat diuji dengan melihat pola plot  $\ln[-\ln(S(t))]$  atau  $-\ln[-\ln(S(t))]$  terhadap waktu *survival* ( $t$ ) untuk setiap variabel prediktor dengan skala kategorik dan grafik *observed versus expected* kurva *survival*. Asumsi *proportional hazard* terpenuhi jika pola plot antar kategori

dalam variabel prediktor membentuk pola yang sejajar untuk plot  $\ln[-\ln(S(t))]$  dan pola plot yang berhimpitan untuk grafik *observed versus expected* kurva *survival* yang dapat dijelaskan pada Gambar 2.10 dan 2.11.



**Gambar 2.10** Ilustrasi Kurva  $-\ln(-\ln S(t))$  (Kleinbaum & Klein, 2012)



**Gambar 2.11** Ilustrasi Grafik *Observed versus Expected* (Kleinbaum & Klein, 2012)

Diilustrasikan bahwa terdapat variabel jenis kelamin (*sex*) laki-laki (*males*) dan perempuan (*females*). Selanjutnya akan dilakukan analisis apakah variabel *sex* memenuhi asumsi *proportional hazard* atau tidak. Berdasarkan Gambar 2.10, terlihat bahwa asumsi *proportional hazard* terpenuhi karena garis yang mewakili data *males* sejajar dengan garis yang mewakili data *females*. Apabila dilihat dari grafik *observed versus expected*, variabel *sex* dikatakan memenuhi asumsi *proportional hazard* karena kurva *survival* pengamatan (*observed*) dan prediksi (*expected*) saling berdekatan (Kleinbaum & Klein, 2012).

b. Uji *Goodness of Fit*

Pengujian asumsi *proportional hazard* dengan metode *goodness of fit* menggunakan residual *Schoenfeld*. Residual *Schoenfeld* terdefinisi pada setiap individu yang mengalami *event* untuk setiap variabel prediktor dalam model

(Kleinbaum & Klein, 2005). Adapun langkah-langkah pengujian asumsi *proportional hazard* menggunakan residual *Schoenfeld* adalah sebagai berikut (Schoenfeld, 1982).

1. Membangun model Cox *proportional hazard* dengan metode *goodness of fit* menggunakan residual *Schoenfeld* untuk setiap variabel prediktor.
2. Membuat variabel rank *survival time* dimana waktu *survival* diurutkan mulai dari individu yang mengalami *event* pertama kali.
3. Menguji korelasi antara variabel yang dihasilkan pada langkah pertama yaitu residual *Schoenfeld* dengan variabel yang dihasilkan pada langkah kedua yaitu *rank survival time*.

Residual *Schoenfeld* dari variabel prediktor ke- $p$  dan individu yang mengalami *event* pada waktu  $t_{(j)}$  didefinisikan sebagai berikut:

$$PR_{pj} = x_{pj} - E(x_{pj} | R(t_{(pj)})) \quad (2.20)$$

Dengan

$$E(x_{pj} | R(t_{(pj)})) = \frac{\sum_{l \in R(t_{(j)})} x_{pj} \exp(\beta^T \mathbf{x}_l)}{\sum_{l \in R(t_{(j)})} \exp(\beta^T \mathbf{x}_l)} \quad (2.21)$$

$PR_{pj}$  : residual *Schoenfeld* untuk variabel ke- $p$  individu yang mengalami *event* pada waktu  $t_{(j)}$

$x_{pj}$  : nilai dari variabel prediktor ke- $p$  dari individu yang mengalami *event* pada waktu  $t_{(j)}$

$E(x_{pj} | R(t_{(pj)}))$  : *conditional expectation*  $x_{pj}$  jika  $R_{t(j)}$  diketahui,

Dalam pengujian korelasi antara residual *Schoenfeld* dengan *rank survival time* masing-masing variabel prediktor digunakan koefisien korelasi pearson.

$$\rho = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left( \sum_{i=1}^n y_i \right)^2}} \quad (2.22)$$

Dengan

$n$  : banyaknya individu

$x_i$  : residual *Schoenfeld* individu ke- $i$

$y_i$  : *rank survival time* individu ke- $i$

Statistik uji korelasi pearson

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \quad (2.23)$$

Tolak  $H_0$ , jika  $|t| > t_{(\alpha/2, n-1)}$

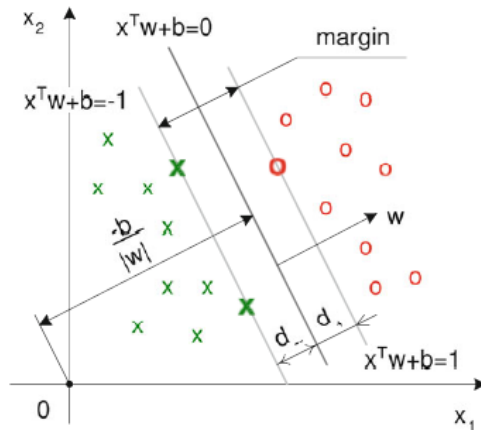
### c. Variabel *Time Dependent*

Pemeriksaan asumsi *proportional hazard* juga dapat dilakukan dengan uji variabel *time dependent*. Variabel *time dependent* adalah variabel prediktor dalam model *Cox proportional hazard* yang diinteraksikan dengan fungsi waktu.

## 2.4 Support Vector Machine

*Support Vector Machine* (SVM) adalah suatu teknik klasifikasi yang relatif baru. Para ilmuwan dan praktisi telah banyak menerapkan teknik ini dalam menyelesaikan masalah-masalah yang nyata dalam kehidupan sehari-hari. Baik dalam masalah *gene expression*, analisis, finansial, cuaca hingga di bidang kedokteran. SVM melakukan suatu teknik untuk menemukan fungsi pemisah (*classifier*) yang bisa memisahkan dua set data dari dua kelas yang berbeda (Vapnik, 1995).

Secara konseptual, SVM adalah mesin linier yang dilengkapi dengan fitur spesial (Gorunescu, 2011), serta teori pembelajaran statistik. Dua sifat khusus dari SVM yaitu (1) mencapai generalisasi yang tinggi dengan memaksimalkan margin, dan (2) mendukung pembelajaran yang efisien dari fungsi nonlinier pada trik kernel sehingga membuat kinerja generalisasinya baik dalam menyelesaikan masalah pengenalan pola (Gorunescu, 2011). Untuk permasalahan klasifikasi SVM mencoba untuk mencari garis pemisah (*hyperplane*) yang optimal.



**Gambar 2.12** Konsep *Hyperplane* pada SVM (Haerdle, *et.al.*, 2014)

Gambar 2.12 menunjukkan sebuah data set yang memiliki dua kelas yaitu kelas  $\{-1\}$  dan  $\{1\}$ . *Hyperplane* terbaik merupakan *hyperplane* yang memiliki margin maksimal yang diperoleh dari alternatif garis pemisah (*discriminant boundaries*). *Margin* adalah jarak antara *hyperplane* dengan titik terdekat dari masing-masing kelas. Bidang pembatas pertama membatasi kelas pertama dan bidang pembatas kedua membatasi kelas kedua sedangkan data yang berada pada bidang pembatas merupakan vektor-vektor yang terdekat dengan *hiperplane* terbaik disebut *support vector* (Nugroho, 2003).

Fungsi klasifikasi  $\mathbf{x}^T \mathbf{w} + b$  berada dalam sebuah keluarga fungsi klasifikasi  $\mathcal{F}$  yang terbentuk yaitu  $\mathbf{x}^T \mathbf{w} + b, \mathbf{w} \in \mathbb{R}^p, b \in \mathbb{R}$ .

Bidang pemisah (*separating hyperplane*):

$$f(x) = \mathbf{x}^T \mathbf{w} + b = 0 \quad (2.24)$$

Fungsi pemisah untuk kedua kelas adalah sebagai berikut:

$$\mathbf{x}_i^T \mathbf{w} + b \geq 1 \text{ untuk } y_i = +1 \quad (2.25)$$

$$\mathbf{x}_i^T \mathbf{w} + b \leq -1 \text{ untuk } y_i = -1 \quad (2.26)$$

Dengan  $\mathbf{w}$  adalah vektor bobot (*weight vector*) yang berukuran  $(p \times 1)$ ,  $b$  adalah posisi bidang relatif terhadap pusat koordinat atau lebih dikenal dengan bias yang bernilai skalar. Pada Gambar 2.12 menunjukkan  $\frac{|b|}{\|\mathbf{w}\|}$  adalah jarak bidang pemisah yang tegak lurus dari titik pusat koordinat dan  $\|\mathbf{w}\|$  adalah jarak euclidian dari  $\mathbf{w}$ . Bidang pembatas pertama membatasi kelas  $(+1)$  sedangkan bidang pembatas

kedua membatasi kelas (-1). *Hyperplane* yang optimal adalah  $\max \frac{2}{\|\mathbf{w}\|}$  atau *equivalent* dengan  $\min \frac{1}{2} \|\mathbf{w}\|^2$ . Dengan menggabungkan kedua konstrain pada persamaan (2.25) dan (2.26) maka dapat direpresentasikan dalam pertidaksamaan sebagai berikut:

$$y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1 \geq 0, \quad i = 1, 2, \dots, n \quad (2.27)$$

Secara matematis, formulasi problem optimasi SVM untuk klasifikasi linier dalam *primal space* adalah

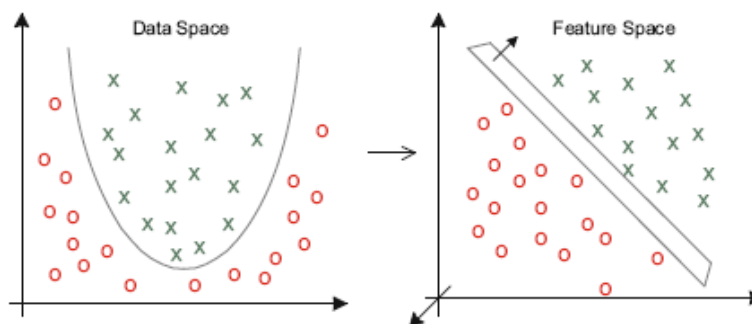
$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.28)$$

dengan fungsi kendala  $y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1 \geq 0, \quad i = 1, 2, \dots, n$ . Secara umum, persoalan optimasi (2.28), akan lebih mudah diselesaikan jika diubah ke dalam formula *Lagrange*. Dengan demikian permasalahan optimasi dengan konstrain dapat dirumuskan menjadi:

$$L_{\text{pri}}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i \{y_i(\mathbf{x}_i^T \mathbf{w} + b) - 1\} \quad (2.29)$$

Dengan  $\alpha_i \geq 0$  (Haerdle, *et al.*, 2011).

Pada dasarnya, SVM merupakan linier *classifier* tetapi seiring berkembangnya penelitian SVM dapat bekerja pada masalah non linier dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi.



**Gambar 2.13** Mapping dari Dua Dimensi Data Space (Kiri) ke Tiga Dimensi Feature Space (Kanan) (Haerdle *et al.*, 2014)

Dalam mencari solusi masalah nonlinier digunakan “*kernel trick*” yaitu menambahkan fungsi kernel ke dalam persamaan SVM. Menurut Hsu *et al.* (2010), terdapat 4 fungsi kernel yaitu

1. Kernel Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (2.30)$$

2. Kernel Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mu \mathbf{x}_i^T \mathbf{x}_j + r)^d, \mu > 0 \quad (2.31)$$

3. Fungsi Kernel Radial Basis Function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (2.32)$$

4. Kernel Eksponensial

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\mu \mathbf{x}_i^T \mathbf{x}_j + r) \quad (2.33)$$

dengan  $\mu$ ,  $r$ ,  $d$  dan  $\sigma$  merupakan parameter kernel dan  $i, j=1,2,\dots,m$ .

Penggunaan kernel bertujuan untuk mengimplementasikan suatu model pada ruang dimesi yang lebih tinggi (*feature space*) sehingga kasus yang *non lineary separable* pada ruang input bisa ditransformasi menjadi *lineary separable* pada *feature space*. Pada kasus nonlinier, masalah optimasi digunakan untuk menemukan fungsi yang paling sesuai pada *feature space*, bukan pada *input space*.

Pemilihan fungsi kernel yang tepat merupakan hal yang sangat penting karena akan menentukan *feature space* dimana fungsi *classifier* akan dicari. Sepanjang fungsi kernelnya sesuai (cocok), SVM akan beroperasi secara benar meskipun tidak tahu pemetaan yang digunakan (Santosa, 2007; Robandi, 2008). Menurut Scholkopf dan Smola (2002), fungsi kernel gaussian RBF memiliki kelebihan yaitu secara otomatis menentukan nilai, lokasi dari *center* dan nilai pembobot dan bisa mencakup nilai rentang tak terhingga. Gaussian RBF juga efektif menghindari *overfitting* dengan memilih nilai yang tepat untuk parameter C dan  $\sigma$  dan RBF baik digunakan ketika tidak ada pengetahuan terdahulu. Menurut Hsu, Chang dan Lin (2004), fungsi kernel yang direkomendasikan untuk diuji pertama kali adalah fungsi kernel RBF karena dapat memetakan hubungan tidak linier RBF



lebih robust terhadap *outlier* karena fungsi kernel RBF berada antara selang  $(-\infty, \infty)$  sedangkan fungsi kernel yang lain memiliki rentang antara (-1 sampai dengan 1).

## 2.5 Least Square Support Vector Machine (LS-SVM)

Suyken dan Vandewalle (1999) mengusulkan sebuah versi *least Squares* untuk algoritma pembelajaran *Support Vector Machine* (SVM) yang disebut *Least Squares Support Vector Machine* (LS-SVM). LS-SVM adalah modifikasi metode SVM standar yang mengarah pada pemecahan linier sistem *Karush Kuhn Tucker* (KKT). Dalam formulasi LS-SVM, perhitungan komputasi dari SVM yang disederhanakan dengan pelaksanaan versi *Least Square* (LS) daripada *inequality constraints* dan fungsi penalti penjumlahan kesalahan kuadrat (*squared error*) sebagaimana digunakan dalam pelatihan jaringan saraf tiruan. Reformulasi ini sangat menyederhanakan masalah dalam memecahkan satu set persamaan linier dari pada pemrograman kuadratik (*quadratic programming*) yang digunakan dalam SVM standar. *Primal problem* pada LS-SVM atau fungsi tujuan dinyatakan pada persamaan (2.34).

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 \quad (2.34)$$

Fungsi kendala  $y_i [\varphi(\mathbf{x}_i)^T \mathbf{w} + b] = 1 - \xi_i; i = 1, 2, \dots, n$

Fungsi kernel  $\Omega_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$

Sehingga didapatkan fungsi Lagrange sebagai berikut:

$$L_{\text{pri}}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i (y_i [\varphi(\mathbf{x}_i)^T \mathbf{w} + b] - 1 + \xi_i) \quad (2.35)$$

Dengan  $\alpha_i$  adalah pengali Lagrange (dapat bernilai positif atau negatif). Dari persamaan fungsi Lagrange untuk kondisi optimal dapat ditunjukkan pada persamaan (2.36).

$$\begin{cases} \frac{\partial L_{pm}(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \varphi(\mathbf{x}_i) \\ \frac{\partial L_{pm}(\mathbf{w}, b, \alpha)}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L_{pm}(\mathbf{w}, b, \alpha)}{\partial \xi_i} = 0 \rightarrow \alpha_i = C \xi_i, i = 1, \dots, n \\ \frac{\partial L_{pm}(\mathbf{w}, b, \alpha)}{\partial \alpha_i} = 0 \rightarrow y_i [\varphi(\mathbf{x}_i)^T \mathbf{w} + b] = 1 - \xi_i, i = 1, \dots, n \end{cases} \quad (2.36)$$

Persamaan (2.36) ditulis sebagai sistem linier sebagai ganti dari *Quadratic Programming* sebagai berikut:

$$\begin{bmatrix} \mathbf{I} & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -y^T \\ 0 & 0 & C\mathbf{I} & -\mathbf{I} \\ Z & y & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ b \\ \xi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{1} \end{bmatrix} \quad (2.37)$$

Dengan

$\mathbf{Z} = [\varphi(\mathbf{x}_1)^T y_1, \dots, \varphi(\mathbf{x}_n)^T y_n]^T$ ,  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ ,  $\mathbf{1} = [1, 1, \dots, 1]^T$ ,  $\xi = [\xi_1, \dots, \xi_n]^T$ ,  
 $\alpha = [\alpha_1, \dots, \alpha_n]^T$ ,  $e = [e_1, \dots, e_n]^T$ ,  $C$  adalah parameter regulasi atau ongkos penalti akibat *misclassification* dan  $\mathbf{I}$  adalah matrik identitas. Setelah mengeliminasi  $\mathbf{w}$  dan  $\xi$  mengikuti sistem linier Karush Kuhn Tucker (KKT) sehingga menghasilkan persamaan (2.38).

$$\begin{bmatrix} 0 & y^T \\ y & \mathbf{Z}\mathbf{Z}^T + C^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (2.38)$$

$$\begin{bmatrix} 0 & y^T \\ y & \Omega + C^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix} \quad (2.39)$$

Kondisi Mercer diaplikasikan ke matriks  $\Omega = \mathbf{Z}\mathbf{Z}^T$  dan kernel *trick* diaplikasikan dalam matriks  $\Omega$

$$\begin{aligned} \Omega_{ij} &= y_i y_j \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \\ &= y_i y_j K(\mathbf{x}_i, \mathbf{x}_j); i = 1, 2, \dots, n \end{aligned} \quad (2.40)$$

## 2.6 Survival SVM (SURSVSM)

Metodologi menggunakan Support Vector Machine (SVM) untuk analisis data *survival*, pertama kali diteliti oleh (Van Belle, *et al.*, 2007 & 2008). Sebagai ganti dari fungsi *hazard*, kesesuaian antara observasi waktu kegagalan dan *output* model dioptimalkan. *Output* model merupakan fungsi prognostik juga disebut sebagai fungsi utilitas dan lebih spesifik dalam penelitian medis disebut prognostik indeks atau fungsi kesehatan dimana  $u : \mathbb{R}^d \rightarrow \mathbb{R}$  didefinisikan pada persamaan (2.41).

$$u(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) \quad (2.41)$$

Dimana  $w$  vektor parameter yang tidak diketahui dan  $\boldsymbol{\varphi}(x)$  merupakan transformasi dari kovariat  $x$ . Pada SVM nonlinear, fungsi transformasi yang digunakan adalah “*Kernel Trick*” (Scholkopf & Smola, 2002). *Kernel Trick* menghitung *scalar product* dalam bentuk sebuah fungsi kernel. Proyeksi  $\boldsymbol{\varphi} : \mathbb{R}^d \rightarrow \mathbb{R}$  memastikan bahwa *inner product* dipresentasikan oleh fungsi kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j) \quad (2.42)$$

Jika sebuah fungsi kernel  $K$  pada persamaan (2.42), ini dapat digunakan tanpa perlu mengetahui fungsi transformasi  $\boldsymbol{\varphi}$  secara eksplisit. Diberikan sebuah kernel  $K$  dan data  $x_1, x_2, x_3, \dots, x_n \in \mathbf{x}$  maka matrik  $K = (K(\mathbf{x}_i, \mathbf{x}_j))_{ij}$  berukuran  $n \times n$  disebut *Gram matrix* untuk data  $x_1, x_2, x_3, \dots, x_n$ . Sebuah syarat cukup dan perlu untuk matrik simetri  $K$  dengan  $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_j, \mathbf{x}_i) = K_{ji}$ , untuk  $K$  definit positif disebut *Mercer’s Theorem* (Mercer, 1909).

$$\sum_{i=1}^n \sum_{j=1}^n \boldsymbol{\varphi}_i \boldsymbol{\varphi}_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0 \quad (2.43)$$

Contoh sederhana pada sebuah *kernel trick* yang menunjukkan bahwa kernel dapat dihitung tanpa perhitungan fungsi *mapping*  $\boldsymbol{\varphi}$  secara eksplisit adalah fungsi pemetaan :

$$\boldsymbol{\varphi}(x_1, x_2) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)^T$$

Sehingga menjadi

$$\mathbf{w}^T \boldsymbol{\phi}(x) = w_1 x_1^2 + \sqrt{2} w_2 x_1 x_2 + w_3 x_2^2$$

dengan dimensi pada *feature space* adalah kuadratik, padahal dimensi asalnya adalah linier. Metode kernel menghindari pembelajaran secara eksplisit *mapping* data ke dalam *feature space* dimensi tinggi, seperti pada contoh berikut.

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w}^T \mathbf{x} + b \\ &= \sum_{i=1}^n \alpha_i \mathbf{x}_i^T \mathbf{x} + b \\ &= \sum_{i=1}^n \alpha_i \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}) + b \\ &= \sum_{i=1}^n \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \end{aligned}$$

Hubungan kernel dengan fungsi *mapping* adalah sebagai berikut:

$$\begin{aligned} \boldsymbol{\phi}(\mathbf{x}_i)^T \boldsymbol{\phi}(\mathbf{x}) &= (x_{i1}^2, \sqrt{2}x_{i1}x_{i2}, x_{i2}^2)(x_1^2, \sqrt{2}x_1x_2, x_2^2)^T \\ &= x_{i1}^2 x_1^2 + 2x_{i1}x_{i2}x_1x_2 + x_{i2}^2 x_2^2 \\ &= (\mathbf{x}_i^T \mathbf{x})^2 \\ &= K(\mathbf{x}_i, \mathbf{x}) \end{aligned}$$

Diasumsikan bahwa  $t_i < t_j$  untuk  $i < j$ . Jika asumsi tersebut dilanggar, maka data diurutkan terlebih dahulu. Dengan mengetahui bahwa  $t_i$  lebih kecil dari  $t_j$ , diharapkan bahwa nilai  $u(x_i)$  juga lebih kecil dari  $u(x_j)$ . Sehingga

$$u(x_j) - u(x_i) \geq 1, \forall i < j \quad (2.44)$$

Model *survival SVM* diimplikasikan dengan fungsi kendala (*constraint*) yang akan mendapatkan margin yang tepat. Fungsi kendala model *survival SVM* ditunjukkan pada persamaan (2.44) di atas. Jika terjadi kesalahan dalam memberi peringkat maka diberi variabel slack yaitu  $\xi_{ij} \geq 0$ . Sehingga formulasi model *survival SVM* dinyatakan pada persamaan (2.45).

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i < j} v_{ij} \xi_{ij} \quad (2.45)$$

$$\text{Fungsi Kendala} \begin{cases} \mathbf{w}^T \boldsymbol{\phi}(x_j) - \mathbf{w}^T \boldsymbol{\phi}(x_i) \geq 1 - \xi_{ij}, \forall i < j \\ \xi_{ij} \geq 0, \forall i < j \end{cases} \quad (2.46)$$

Dengan  $\gamma \geq 0$  dan  $w$  didefinisikan pada persamaan (2.41).

## 2.7 Additive Survival Least Square SVM (A-SURLSSVM)

Additive survival Least square SVM (A-SURLSSVM) merupakan model modifikasi dari SURLSSVM (Van Belle *et al.*, 2010). Fungsi tujuan model A-SURLSSVM adalah sebagai berikut:

$$\min_{w, \xi, b, \chi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 + \frac{1}{2} \mu \sum_{j=1}^n \chi_j^2 \quad (2.47)$$

$$\text{Fungsi kendala} \begin{cases} \mathbf{w}^T \varphi(x_j) - \mathbf{w}^T \varphi(x_i) = 1 + \xi_{ij}, \quad \forall i, j = 1, 2, \dots, n \\ \delta_j t_j = \delta_j (\mathbf{w}^T \varphi(x_j) + b) - \chi_j, \quad \forall j = 1, 2, \dots, n \end{cases} \quad (2.48)$$

Dengan  $\gamma \geq 0$  dan  $\mu \geq 0$ . Berdasarkan fungsi tujuan dan fungsi kendala pada persamaan (2.47) dan (2.48) maka persamaan Lagrange dari model A-SURLSSVM adalah sebagai berikut:

$$\begin{aligned} L(\mathbf{w}, \xi, \chi, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 + \frac{1}{2} \mu \sum_{j=1}^n \chi_j^2 \\ &\quad - \sum_{i,j=1}^n \alpha_{ij} (\mathbf{w}^T (\varphi(x_j) - \varphi(x_i)) - 1 - \xi_{ij}) \\ &\quad - \sum_{j=1}^n \beta_j (\delta_j (-t_j + \mathbf{w}^T \varphi(x_j) + b) - \chi_j) \end{aligned} \quad (2.49)$$

Kondisi Karush Kuhn Tucker (KKT) untuk kondisi optimal didapatkan dari turunan pada persamaan (2.49) (Boyd, S., & Vandenberghe, L., 2004). Setelah mengeliminasi  $w, \xi$  dan  $\chi$  dari kondisi KKT maka didapatkan persamaan (2.50).

$$\begin{bmatrix} \mathbf{D} \mathbf{K} \mathbf{D}^T + \mathbf{I}_{n_c} / \gamma & \mathbf{D} \mathbf{K} \mathbf{R} & 0 \\ \mathbf{R} \mathbf{K} \mathbf{D}^T & \mathbf{R} \mathbf{K} \mathbf{R} + \mathbf{I}_n / \mu & \boldsymbol{\delta} \\ 0 & \boldsymbol{\delta}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{R} \mathbf{t} \\ 0 \end{bmatrix} \quad (2.50)$$

Dengan  $\alpha$  dan  $\beta$  adalah Lagrange Multiplier yang berhubungan dengan konstrain,  $t = [t_1, t_2, \dots, t_n]^T$ ,  $\delta = [\delta_1, \delta_2, \dots, \delta_n]^T$  dan  $R \in \mathbb{R}^{n \times n}$  adalah diagonal matriks  $R = \text{diag}(\delta)$  Untuk mendapatkan prognostik indeks digunakan persamaan berikut.

$$\begin{aligned}
u(x)^* &= w^T \boldsymbol{\varphi}(x^*) + b \\
&= \left( \sum_{i,j} \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T \boldsymbol{\varphi}(x^*) + b \\
&= (\boldsymbol{\alpha}^T \mathbf{D} + \boldsymbol{\beta}^T \mathbf{R}) K_n(\mathbf{x}^*) + b
\end{aligned} \tag{2.51}$$

## 2.8 Kriteria Kebaikan Model

Banyak ukuran yang digunakan untuk membandingkan suatu model dengan model lainnya atau untuk memilih model terbaik.

### a. Concordance Index

*Concordance Index* (C-index) yang diperkenalkan oleh Harrell (1984) digunakan untuk mengukur *Concordance* antara fungsi utilitas dan waktu kegagalan observasi menggunakan observasi tersensor dan tidak tersensor. *Concordance* yang empiris antara waktu terjadinya *event* ( $t_i$ ) dan prognostik indeks  $u_i$  berdasarkan dataset  $D = \{(x_i, t_i, \delta_i)\}_{i=1}^n$  didefinisikan pada persamaan (2.52).

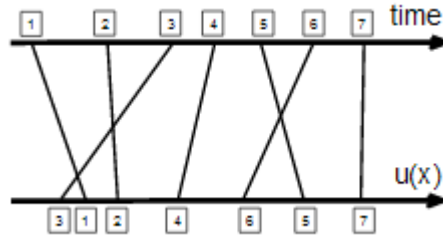
$$c_{ij}(u) = \frac{\sum_{i=1}^n \sum_{j>i}^n v_{ij} I((u(x_j) - u(x_i))(t_j - t_i) > 0)}{\sum_{i=1}^n \sum_{j>i}^n v_{ij}} \tag{2.52}$$

dimana  $((u(x_j) - u(x_i))(t_j - t_i) > 0)$  merupakan fungsi indikator dan  $v_{ij}$  dirumuskan sebagai berikut:

$$v_{ij} = \begin{cases} 1, & (t_i < t_j \text{ dan } \delta_i = 1) \text{ atau } (t_j < t_i \text{ dan } \delta_j = 1) \\ 0, & \text{untuk yang lainnya} \end{cases}$$

dengan  $(x_i, t_i, \delta_i)$  dan  $(x_j, t_j, \delta_j)$  akan dibandingkan ketika peringkat dalam domain waktu diketahui. Sebagai contoh perbandingan dua data yang tidak tersensor atau perbandingan sebuah data yang tidak tersensor dengan sebuah data tersensor kanan. Pada perumusan C-indeks diasumsikan bahwa  $t_i < t_j$  untuk  $i < j$ . Jika asumsi tersebut dilanggar, maka data diurutkan terlebih dahulu. Dengan mengetahui bahwa  $t_i$  lebih kecil dari  $t_j$ , diharapkan bahwa

nilai  $u(x_i)$  juga lebih kecil dari  $u(x_j)$ . Contoh sederhana cara perhitungan c-indeks secara manual.



**Gambar 2.14** Ilustrasi Perhitungan C-Indeks (Van Bell *et al.*, 2011)

Misalkan kasus yang terjadi pada Gambar 2.14, dimana terdapat 7 pasien yang memiliki *survival time* dan prognostik indeks yang telah diperingkat. Berdasarkan Gambar 2.14, dapat diketahui bahwa perhitungan C-indeks adalah sebagai berikut:

**Tabel 2.1** Ilustrasi Perhitungan C-Indeks

<i>Concordance</i>	<i>Ranking</i>	Jumlah
	1 2 3 4 5 6	
2	1	1
3	0 0	0
4	1 1 1	3
5	1 1 1 1	4
6	1 1 1 1 0	4
7	1 1 1 1 1 1	6
Jumlah	5 4 4 3 1 1	c-indeks 18/21

Sumber : Van Bell *et al.* (2011)

Berdasarkan ilustrasi pada Tabel 2.1, terdapat 3 *misranking*. Sebagai contoh peringkat dari prognostik indeks yang *survival timenya* peringkat 3 adalah 1 atau yang memiliki prognostik indeks paling kecil. Itu artinya  $t_i$  lebih kecil dari  $t_j$ , diharapkan bahwa nilai  $u(x_i)$  juga lebih kecil dari  $u(x_j)$  dilanggar (*misranking*). Dari 21 kemungkinan pasangan peringkat  $(t_i, u(x_i))$  dan  $(t_j, u(x_j))$  ada 18 yang tidak *misranking*. Sehingga nilai C-indeks pada ilustrasi di atas adalah  $\frac{18}{21}$ . Semakin tinggi nilai C-indeks maka performansi dari metode yang digunakan semakin bagus (Mahjub, Hossein *et al.*, 2016).

b. Logrank

Setelah mendapatkan nilai prognostik indeks untuk masing-masing metode, nilai prognostik indeks tersebut dikelompokkan menjadi dua kategori yaitu kelompok beresiko tinggi (*high*) dan kelompok beresiko rendah (*low*). Individu yang memiliki prognostik indeks di atas median, masuk ke dalam kelompok beresiko tinggi untuk terjadi *failure* dan sebaliknya (Van Belle, *et.al.*, 2010).

Hipotesis dari uji Logrank untuk dua grup atau lebih adalah sebagai berikut:

$H_0$  : tidak ada perbedaan kelompok beresiko tinggi (*high*) dan kelompok beresiko rendah (*low*).

$H_1$  : ada perbedaan kelompok beresiko tinggi (*high*) dan kelompok beresiko rendah (*low*).

Statistik uji yang digunakan adalah  $\chi^2$ . Semakin tinggi nilai  $\chi^2$  maka performansi dari metode yang digunakan semakin bagus karena dapat membedakan kelompok beresiko tinggi (*high*) dan kelompok beresiko rendah (*low*) (Mahjub, Hossein *et al.*, 2016).

c. *Hazard Ratio*

Ukuran lain yang dapat digunakan untuk mengetahui metode yang memiliki performansi terbaik adalah *hazard ratio*. Dengan cara mengelompokkan pasien beresiko tinggi (*high*) dan beresiko rendah (*low*) berdasarkan mean. Selanjutnya mengestimasi nilai tersebut sebagai *covariates* dari model Cox. Selanjutnya menghitung *Hazard Ratio* dari model Cox yang telah terbentuk (Van Belle, *et.al.*, 2010). Semakin tinggi nilai *Hazard Ratio* maka performansi dari metode yang digunakan semakin bagus (Mahjub, Hossein *et al.*, 2016).

## 2.9 Pembangkitan Data *Survival Time*

Studi simulasi merupakan alat statistik yang penting untuk menyelidiki performansi, sifat dan kecukupan suatu model statistik. Salah satu model statistik yang paling penting dan sering digunakan dalam penelitian medis adalah model Cox PHM. Pembangkitan data simulasi dengan model Cox PHM telah dilakukan.



Model Cox PHM dapat menggambarkan hubungan *hazard* dan *survival time* yang berguna untuk studi simulasi (Bender, 2005).

Fungsi *survival* dari Cox PHM dinyatakan sebagai berikut:

$$S(t | \mathbf{x}) = \exp(-H_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x})) \quad (2.53)$$

Dengan

$H_0(t) = \int_0^t h_0(u) du$  adalah kumulatif bazeline hazard. Sehingga fungsi distribusi dari

Cox PHM adalah

$$\begin{aligned} F(t | \mathbf{x}) &= 1 - S(t | \mathbf{x}) \\ &= 1 - \exp(-H_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x})) \end{aligned} \quad (2.54)$$

Dengan  $U \sim U_{[0,1]}$  yang merupakan representasi dari *survival function*. Sehingga dengan menggunakan persamaan (2.54) waktu *survival* ( $T$ ) untuk model berbasis Cox adalah

$$U = \exp(-H_0(T) \exp(\boldsymbol{\beta}^T \mathbf{x})) \quad (2.55)$$

$$T = H_0^{-1}(T) [-\log(U) \exp(\boldsymbol{\beta}^T \mathbf{x})] \quad (2.56)$$

Pembangkitan data *survival time* berdistribusi eksponensial berbasis Cox juga telah dilakukan. Apabila  $T$  melambangkan waktu *survival* berdistribusi eksponensial dan mempunyai fungsi distribusi peluang  $f(t)$ , maka fungsi kepadatan peluang dapat dinyatakan sebagai berikut:

$$f(t) = \lambda \exp(-\lambda t) \quad (2.57)$$

Fungsi *survival*  $S(t)$  yang didefinisikan sebagai probabilitas dengan formula sebagai berikut:

$$S(t) = \exp(-\lambda t) \quad (2.58)$$

Fungsi *hazard* merupakan suatu laju kegagalan atau *failure* sesaat dengan asumsi bahwa suatu objek mencapai kejadian khusus sampai waktu ke- $t$ , dengan syarat telah bertahan sampai waktu tersebut. Fungsi *hazard* untuk distribusi eksponensial dinyatakan pada persamaan (2.59).

$$h(t) = \frac{f(t)}{S(t)} = \lambda \quad (2.59)$$

Apabila  $T$  melambangkan waktu *survival* berdistribusi eksponensial, dengan  $\lambda > 0$  mempunyai fungsi hazard  $t \geq 0$ , maka kumulatif fungsi hazard adalah

$$H(t) = \int_0^t h(u) du = \lambda t \quad (2.60)$$

Dengan memasukkan persamaan (2.60) ke dalam persamaan (2.55) maka didapatkan persamaan (2.61)

$$\begin{aligned} U &= \exp(-\lambda T \exp(\boldsymbol{\beta}^T \mathbf{x})) \\ \log U &= \log(\exp(-\lambda T \exp(\boldsymbol{\beta}^T \mathbf{x}))) \\ \log U &= (-\lambda T \exp(\boldsymbol{\beta}^T \mathbf{x})) \\ T &= -\frac{\log U}{\lambda \exp(\boldsymbol{\beta}^T \mathbf{x})} \end{aligned} \quad (2.61)$$

Waktu *survival* untuk model eksponensial berbasis Cox pada individu ke- $i$  dapat ditulis sebagai berikut:

$$T_i = -\frac{\log(U_i)}{\lambda \exp(\boldsymbol{\beta}^T \mathbf{x}_i)} \quad (2.62)$$

Sehingga dengan menggunakan persamaan (2.62) waktu *survival* untuk model eksponensial berbasis Cox dapat dibangkitkan untuk studi simulasi,  $T \sim \text{Exp}(\lambda)$ . Sedangkan untuk data yang waktu *survival* terdapat interaksi pada individu ke- $i$  dapat dibangkitkan dengan persamaan sebagai berikut (Kriner, 2007).

$$T_i = -\frac{\log(U_i)}{\lambda \exp(f(\mathbf{x}_i))} \quad (2.63)$$

Dengan  $U \sim U[0,1]$  yang merupakan representasi dari *survival function*.

Sebagai contoh akan dibangkitkan data simulasi dengan  $n = 1000$  dan *covariate* dibangkitkan dengan distribusi sebagai berikut:

$$\begin{aligned} x_1 &\sim U[30, 75] & x_4 &\sim U[0, 10] \\ x_2 &\sim \text{BIN}(n, 0.5) & x_5 &\sim \text{BIN}(n, 0.7) \\ x_3 &\sim N(0, 1) & U &\sim U[0, 1] \\ \lambda &= 0,5 \end{aligned}$$

dan waktu *survival* dari contoh di atas dapat dibangkitkan dengan persamaan sebagai berikut:

$$t_i = \frac{-\log(U_i)}{0,5 \exp(0,04 \left(\frac{x_{1i}}{10}\right)^2 - 6x_{2i} + 0,1x_{1i}x_{2i})}, i = 1, 2, \dots, n \quad (2.64)$$

## 2.10 *Features Selection* dengan Menggunakan *Backward Elimination*

*Features selection* adalah suatu kegiatan yang umumnya bisa dilakukan secara preprocessing dan bertujuan untuk memilih feature yang berpengaruh dan mengesampingkan feature yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data.

Tujuan penggunaan *features selection*, diantaranya:

1. Dapat mempercepat algoritma machine learning
2. Mengurangi kompleksitas model dan memudahkan untuk interpretasi
3. Meningkatkan keakuratan/performansi model jika tepat dalam pemilihan *features*/variabel
4. Mengurangi *overfitting*/varians yang tinggi

Metode *backward elimination* adalah salah satu metode dalam *features selection*.

Langkah-langkah metode *backward elimination* adalah sebagai berikut:

1. Memasukkan semua variabel ke dalam model
2. Menghilangkan/mengeliminasi satu variabel yang memberikan kontribusi paling kecil terhadap model dalam penelitian ini menggunakan C-index sebagai ukuran performansi
3. Mengulangi langkah 2, hingga tidak ada nilai C-index yang naik (Chandrashekar dan Sahin, 2014).

## 2.11 **Kanker Serviks**

Kanker serviks (kanker leher rahim) merupakan kanker yang paling umum kedua di dunia pada perempuan, dengan perkiraan 471.000 kasus baru (233.000 meninggal dunia) pada tahun 2000. Dari data tersebut, lebih dari 80% penderita berasal dari negara berkembang di Asia Selatan, Asia Tenggara, Sub-Saharan Afrika, Amerika Tengah, dan Amerika Selatan (WHO, 2005). Di Indonesia kanker serviks menempati peringkat kedua setelah kanker payudara dari segi jumlah

penderita kanker pada perempuan. Namun, sebagai penyebab kematian masih menempati peringkat pertama (Ocviyanti & Handoko, 2013). Ketidaktahuan para perempuan akan ancaman kanker serviks turut membantu banyaknya perempuan yang meninggal akibat penyakit ini (Murniati & Lisuwarni, 2014). Berdasarkan data WHO pada tahun 2008, terdapat 530.202 kasus baru kanker serviks di seluruh dunia. Dengan jumlah itu berarti diperkirakan akan didapatkan 1 kasus baru kanker serviks setiap menitnya di dunia. Secara keseluruhan diperkirakan insidensi kanker serviks di seluruh dunia adalah sebesar 16,2 per 100.000 penduduk (Ocviyanti & Handoko, 2013).

Tingkat keparahan pasien kanker serviks dapat dilihat dari stadium. Menurut FIGO (*The International Federation of Gynecology and Obstetrics*), berikut adalah tanda-tanda dan gejala yang dialami pasien kanker serviks pada setiap stadium.

**Tabel 2.2** FIGO (*The International Federation of Gynecology and Obstetrics*) Stadium untuk Kanker Serviks

Stadium	Deskripsi
0	Karsinoma setempat : terlihat perubahan seperti kanker, tetapi tanpa hasil temuan bahwa kanker telah menyebar ke jaringan sekelilingnya (stroma)
I	<p>Kanker terbatas pada serviks</p> <p>Ia Kanker yang menyebar dengan penyebaran ke dalam jaringan penopang tidak lebih dari 5mm dalamnya dan 7mm lebarnya.</p> <p>Ia1 Penyebaran ke dalam jaringan penopang dalamnya 3 mm atau kurang.</p> <p>Ia2 ukuran penyebaran ke dalam jaringan penopang lebih dari 3 mm dan kurang dari 5mm.</p> <p>Ib lesi di serviks terlihat dalam pengamatan atau menunjukkan penyebaran ke dalam jaringan penopang lebih luas daripada lesi 1a.</p> <p>Ib1 tumor kurang dari 4 cm.</p> <p>Ib2 tumor lebih dari 4 cm.</p>

**Tabel 2.3** FIGO (*The International Federation of Gynecology and Obstetrics*) Stadium untuk Kanker Serviks (Lanjutan)

Stadium	Deskripsi
II	Kanker meluas ke rahim IIa Tidak ada tanda yang jelas menyebar ke jaringan lunak yang bersebelahan (parametrium). IIb Parametrium jelas tertular.
III	Kanker meluas sampai ke vagina atau menyebabkan gangguan fungsi ginjal IIIa Tumor sampai ke vagina. IIIb Tumor meluas ke dinding oinggul dan atau mempengaruhi kemampuan ginjal untuk mengeluarkan urine (hidronefrosis).
IV	Karsinoma telah meluas di luar pinggul atau menyerang kandung kemih atau rektum. Iva Tumor menyebar ke kandung kemih atau rektum. Ivb Tumor ditemukan di luar pinggul (paru-paru dan hati)

Sumber : WHO (2005)

### 2.11.1 Faktor Penyebab Kanker Serviks

Banyak faktor berkaitan dengan perkembangan kanker serviks, antara lain :

1. Terinfeksi *human papilloma virus* (HPV)

Menurut *American cancer society* (2014), penyebab utama kanker serviks adalah karena infeksi virus yang bernama *human papiloma virus* (HPV). Jenis virus HPV yang paling fatal adalah HPV tipe 16 dan 18. Cara penyebaran virus ini adalah melalui hubungan seksual terutama dengan berganti-ganti pasangan. Tak hanya melalui cairan, virus ini bisa berpindah melalui sentuhan kulit.

2. Terinfeksi *human immunodeficiency virus* (HIV)

*Human immunodeficiency virus* (HIV) adalah virus penyebab AIDS yang merusak sistem kekebalan tubuh dan meningkatkan risiko terinfeksi HPV

pada perempuan. Hal ini menjelaskan bahwa perempuan dengan penyakit AIDS mengalami peningkatan risiko untuk kanker serviks (Society, 2014).

3. Terinfeksi *chlamydia*

*Chlamydia* pada umumnya adalah jenis bakteri yang dapat menginfeksi sistem reproduksi yang menyebar melalui hubungan seksual. Beberapa penelitian menyebutkan perempuan yang hasil tes darahnya menunjukkan tanda-tanda dari terinfeksi *chlamydia* memiliki risiko yang tinggi untuk kanker serviks dibandingkan dengan perempuan dengan hasil tes normal (Society, 2014).

4. Usia

Menurut McCormick dan Giuntoli (2011) rata-rata perempuan dengan kanker serviks adalah sekitar 50 tahun. Namun, satu dari lima perempuan dengan kanker serviks setidaknya berusia 75 tahun. Dengan bertambahnya usia perempuan, maka risiko untuk kanker serviks mengalami peningkatan. Perempuan dengan usia lanjut mungkin tidak menjalani penapisan yang cukup untuk kanker serviks, dan bila kanker ditemukan, terlalu sering diabaikan atau tidak cukup diobati. Akibatnya, perempuan dengan usia lanjut sering memiliki kanker yang lebih lanjut dan hasil yang lebih buruk dari pasien yang lebih muda.

5. Penggunaan pil KB

Satu faktor risiko yang telah diteliti adalah penggunaan pil KB. Analisis literatur yang baru-baru ini menyimpulkan bahwa mungkin terdapat kenaikan risiko dua kali lipat berkembangnya kanker serviks diantara perempuan yang menggunakan pil KB selama lebih dari 5 tahun (Society, 2014).

6. Merokok

Perempuan yang merokok mengalami peningkatan risiko dua kali lipat untuk kanker serviks dibandingkan dengan perempuan yang tidak merokok. Peneliti percaya bahwa tembakau merusak DNA dari sel serviks dan berkontribusi untuk mengembangkan kanker serviks. Merokok juga membuat sistem kekebalan tubuh berkurang untuk melawan infeksi HPV (McCormick & Giuntoli, 2011).

7. Kelebihan berat badan (*overweight*)  
Perempuan yang memiliki kelebihan berat badan memungkinkan untuk mengembangkan *adenocarcinoma* (kanker tipe kelenjar yang muncul dari bagian-bagian berbeda dari badan) pada serviks (Society, 2014).
8. Diet rendah sayur-sayuran dan buah-buahan  
Perempuan yang berdiet tanpa mengonsumsi sayur-sayuran dan buah-buahan yang cukup akan meningkatkan risiko untuk kanker serviks (Society, 2014).
9. Kemiskinan  
Kemiskinan juga menjadi faktor untuk kanker serviks. Banyak wanita dengan pendapatan yang kurang, memiliki akses terbatas sehingga tidak dapat melakukan *papanicolaou (Pap) smear* (Society, 2014). *Pap smear* dirancang untuk menemukan bukti infeksi persiten (infeksi yang tidak akan hilang) pada serviks (McCormick & Giuntoli, 2011). Selain itu, menurut Dizon, *et al.*(2011) perempuan dengan pendidikan di bawah sekolah menengah memiliki risiko untuk kanker serviks.
10. Memiliki riwayat keluarga menderita kanker serviks  
Perempuan yang memiliki ibu atau saudara perempuan dengan kanker serviks, maka akan berisiko 2 sampai 3 kali lebih tinggi dari pada tidak memiliki keluarga dengan kanker serviks (Society, 2014).
11. Aktifitas seksual  
Para peneliti telah mengenali sejumlah faktor lain yang penting bagi perkembangan kanker serviks, diantaranya adalah indikator dari aktivitas seksual, termasuk jumlah pasangan seksual, umur saat melakukan hubungan seksual pertama kali, berapa kali telah hamil, dan sejarah penyakit menular seksual. Hasil penelitian menunjukkan bahwa semakin tinggi aktivitas seseorang, semakin besar kemungkinan dia terinfeksi HPV (Dizon, *et al.*, 2011). Menurut American Cancer Society (2014) perempuan yang telah hamil 3 kali atau lebih memiliki peningkatan risiko untuk menderita kanker serviks.

### 2.11.2 Pencegahan Kanker Serviks

Menurut Dalimartha (2004) pencegahan kanker serviks yang utama adalah sebagai berikut:

- a. Tidak melakukan hubungan seksual di usia remaja

Hubungan seks idealnya dilakukan setelah seorang perempuan benar-benar matang. Ukuran kematangan bukan hanya dilihat dari ia sudah menstruasi atau belum, tetapi juga tergantung pada kematangan sel-sel mukosa yang terdapat di selaput kulit bagian dalam rongga tubuh. Umumnya sel-sel mukosa baru matang setelah perempuan berusia 20 tahun ke atas. Terutama untuk perempuan yang masih di bawah 16 tahun memiliki resiko yang sangat tinggi terkena kanker serviks bila telah melakukan hubungan seks.

- b. Tidak berganti-ganti pasangan

Risiko terkena kanker serviks lebih tinggi pada perempuan yang berganti ganti pasangan seks dari pada dengan yang tidak. Hal ini terkait dengan kemungkinan tertularnya penyakit kelamin, salah satunya *Human Papiloma Virus* (HPV).

- c. Melakukan vaksinasi HPV

Vaksinasi dapat dilakukan sebelum remaja. Bisa diberikan kepada wanita usia 12-14 tahun, melalui suntikan sebanyak tiga kali berturut-turut tiap 2 bulan seklai dan dilakukan pengulangan satu kali lagi pada sepuluh tahun kemudian. Hal ini dilakukan agar terhindar dari kanker yang mematikan ini. Untuk itu, telah dikembangkan vaksin HPV yang dapat memberikan manfaat yang besar dalam pencegahan penyakit ini.



*(halaman ini sengaja dikosongkan)*

## **BAB 3**

### **METODOLOGI PENELITIAN**

Metodologi penelitian membahas mengenai kajian teoritis, kerangka konsep, sumber data, variabel penelitian, struktur data, dan tahapan penelitian atau langkah analisis beserta diagram alir. Masing-masing akan dijelaskan pada subbab berikut.

#### **3.1 Kajian Teoritis Metode A-SURLSSVM**

Pada subbab ini akan diuraikan pembahasan tentang kajian teoritis metode A-SURLSSVM.

1. Menyelesaikan permasalahan kuadrat dengan membangun fungsi Lagrange dari fungsi tujuan dan fungsi kendala pada persamaan (2.47) dan (2.48).
2. Menurunkan fungsi Lagrange terhadap  $w, \xi, \chi, b, \alpha$  dan  $\beta$  pada persamaan (2.49) pada kondisi Karush Kuhn Tucker (KKT).
3. Mengeliminasi  $w, \xi$ , dan  $\chi$  dengan cara mensubstitusi hasil turunan fungsi Lagrange sehingga menjadi 3 persamaan linier pada persamaan (2.50).
4. Mengoptimalkan parameter kernel RBF (*Radial Basis Function*) yaitu  $\sigma$  serta nilai  $\gamma$  dan  $\mu$  pada A-SURLSSVM. Parameter model akan menemukan nilai parameter yang memberikan performansi yang paling tinggi.
5. Mendapatkan nilai  $\alpha$  dan  $b$  berdasarkan langkah 3 dan 4.
6. Menghitung nilai prognostik indeks dengan menggunakan formula pada persamaan (2.51)

#### **3.2. Kerangka Konsep**

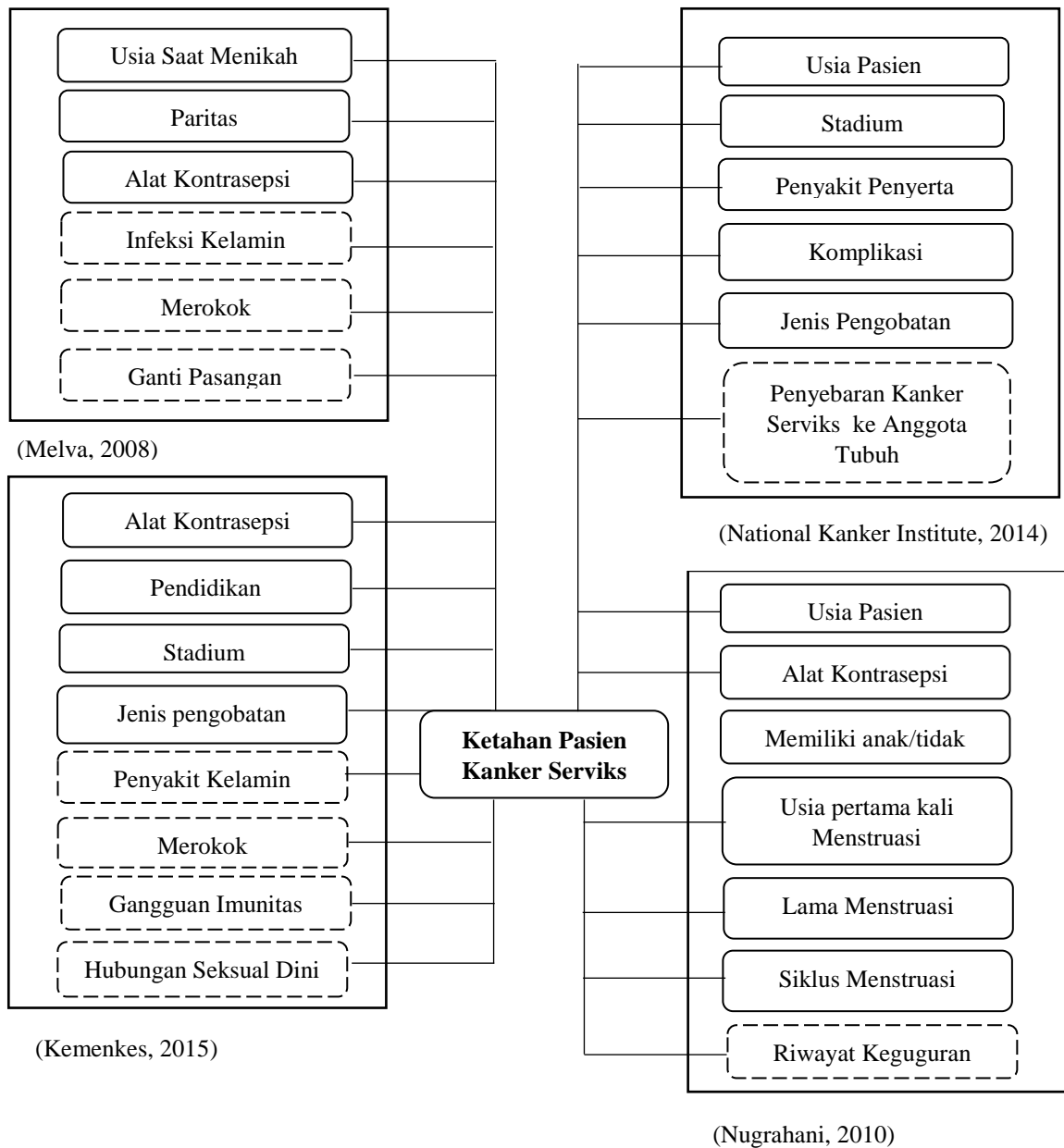
Menurut Kemenkes (2015) beberapa faktor resiko kanker serviks adalah hubungan seksual dini, sosial ekonomi rendah (pendidikan), merokok, penggunaan pil KB (Keluarga Berencana), penyakit yang ditularkan secara seksual dan gangguan imunitas. Angka kematian penderita kanker serviks berkaitan dengan stadium. Stadium kanker serviks didasarkan atas pemeriksaan klinik, stadium klinik

tidak berubah bila kemudian terdapat penemuan baru. Beberapa klasifikasi stadium yaitu 0, IA, IB, IIA, IIB, IIIA, IIIB, IV A, dan IV B. Jenis pengobatan yang diberikan disesuaikan dengan stadium penyakit yaitu operasi, radiasi dan kemoterapi.

Melva (2008) melakukan penelitian tentang faktor-faktor yang mempengaruhi kanker serviks (leher rahim). Dalam penelitian tersebut variabel yang digunakan adalah usia pertama kali menikah, paritas (frekuensi melahirkan), ganti pasangan, penyakit kelamin/infeksi, penggunaan alat kontrasepsi dan status merokok. Setelah dilakukan analisis statistik didapatkan hasil bahwa faktor-faktor yang mempengaruhi kanker serviks adalah usia pertama kali menikah, paritas (frekuensi melahirkan), ganti pasangan, dan penyakit kelamin/infeksi.

Sedangkan menurut *National Cancer Institute* (2014) mengatakan bahwa ketahanan hidup kanker serviks dipengaruhi oleh beberapa faktor diantaranya usia, stadium klinik, kondisi umum sebelum pengobatan, jenis pengobatan dan penyebaran kanker serviks ke anggota tubuh. Selain itu, dalam penelitian ini juga menggunakan beberapa variabel seperti lama menstruasi, siklus menstruasi, serta usia pertama kali menstruasi juga diduga mempengaruhi ketahanan pasien kanker serviks (Nugrahai, 2010).

Oleh karena itu, pada penelitian ini menggunakan beberapa faktor yang mempengaruhi ketahanan hidup pasien kanker serviks di RSUD dr. Soetomo Surabaya dengan variabel penelitian yang meliputi variabel usia pasien, stadium, komplikasi, jenis pengobatan, usia saat menikah, frekuensi melahirkan, usia pertama kali menstruasi, lama menstruasi, siklus menstruasi, penggunaan alat kontrasepsi, dan pendidikan yang digunakan seperti yang tergambar pada Gambar 3.1



Keterangan :

□ : Variabel yang diteliti

□ (dashed) : Variabel yang tidak diteliti

**Gambar 3.1** Kerangka Konsep Penelitian

### 3.3 Aplikasi Metode A-SURLSSVM

Pada subbab ini akan diuraikan tentang sumber data dan variabel penelitian serta aplikasi metode A-SURLSSVM.

#### 3.3.1 Sumber Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini terdiri dari dua tipe data yaitu data simulasi dan data riil. Data simulasi bertujuan untuk menunjukkan bagaimana performansi metode A-SURLSSVM dengan Cox PHM. Selain itu pada data simulasi juga dapat ditentukan banyaknya data yang akan dibangkitkan sehingga mendukung untuk perbandingan performansi kedua metode tersebut. Disisi lain *pattern* data pada kehidupan nyata seringkali tidak sama dengan *pattern* data simulasi, oleh karena itu juga digunakan data riil yaitu data pasien kanker serviks yang pernah menjalani perawatan/pengobatan di RSUD Dr. Soetomo Surabaya. Data pasien kanker serviks diperoleh dari data rekam medik pasien kanker serviks tahun 2013 hingga 2016.

Data simulasi yang digunakan terdiri dari variabel respon berupa *survival time*, status tersensor dan variabel prediktor (*covariate*). Berikut variabel yang digunakan untuk data simulasi.

**Tabel 3.1** Variabel Data Simulasi

Tipe Data <i>Survival Time</i> ( $y$ )	Jumlah Data ( $n$ )	Banyak <i>Covariate</i> ( $p$ )
Tidak terdapat interaksi (Linier dalam <i>covariate</i> )	100	6, 8, 15, & 17
	1000	6, 8, 15, & 17
	5000	6, 8, 15, & 17
Terdapat interaksi (Tidak Linier dalam <i>covariate</i> )	100	6, 8, 15, & 17
	1000	6, 8, 15, & 17
	5000	6, 8, 15, & 17

Berdasarkan Tabel 3.1 variabel respon yaitu *survival time* terdiri dari dua tipe data. *Survival time* dibangkitkan dengan distribusi eksponensial berbasis Cox PHM dengan menggunakan persamaan (2.62) yang berdistribusi eksponensial, yaitu  $y \sim Exp(1)$ . Data dengan *survival time* dibangkitkan yang melibatkan interaksi antara dua *covariate* menggunakan persamaan (2.63) dan diilustrasikan

seperti pada persamaan (2.64). Untuk data dengan *covariate* 6 dan 8, *survival time* dibangkitkan dengan menambahkan interaksi  $0.03*X_1*X_2+0.09*X_2*X_5$ . Untuk *covariate* 15 dan 17, *survival time* dibangkitkan dengan menambahkan interaksi  $0.03*X_1*X_2 + 0.09*X_2*X_{11}$ . Interaksi tersebut hanya digunakan saat membangkitkan *survival time*, namun tidak digunakan sebagai input pada saat pemodelan. Untuk status tersensor, 0 untuk data yang tersensor dan 1 untuk data yang tidak tersensor yang dibangkitkan dari distribusi binomial. Persentase sensing yang digunakan adalah 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, dan 90%. Dalam penelitian ini, banyaknya *covariate* yang digunakan terdiri dari empat skenario yaitu dengan 6,8,15, dan 17. *Covariate* dibangkitkan melalui berbagai distribusi meliputi distribusi diskrit dan kontinu diantara menggunakan distribusi binomial, multinomial, dan normal. Selanjutnya masing-masing skenario disimulasikan dengan jumlah data yang berbeda-beda yaitu 100,1000, dan 5000. Berikut adalah penjelasan untuk masing-masing skenario.

**Tabel 3.2** Skenario untuk Analisis Data Simulasi

Jumlah Variabel	Distribusi	Nilai Parameter
$p = 6$	$X_1 \sim BIN(1, 0.4)$	$\beta_1 = 0.01 \quad \beta_{4(1)} = 0.0015$
	$X_2 \sim BIN(1, 0.3)$	$\beta_2 = -0.15 \quad \beta_{4(2)} = -0.075$
	$X_3 \sim Mult(4, 0.5, 0.2, 0.2, 0.1)$	$\beta_{3(1)} = -0.07 \quad \beta_5 = -0.085$
	$X_4 \sim Mult(3, 0.4, 0.4, 0.2)$	$\beta_{3(2)} = 0.035 \quad \beta_6 = -0.02$
	$X_5 \sim N(40, 2)$	$\beta_{3(3)} = -0.08$
	$X_6 \sim N(20, 3)$	
$p = 8$	$X_1 \sim BIN(1, 0.4)$	$\beta_1 = 0.01 \quad \beta_{4(2)} = -0.075$
	$X_2 \sim BIN(1, 0.3)$	$\beta_2 = -0.15 \quad \beta_5 = 0.085$
	$X_3 \sim Mult(4, 0.5, 0.3, 0.1, 0.1)$	$\beta_{3(1)} = -0.07 \quad \beta_6 = -0.02$
	$X_4 \sim Mult(3, 0.4, 0.4, 0.2)$	$\beta_{3(2)} = 0.035 \quad \beta_7 = 0$
	$X_5 \sim N(40, 2)$	$\beta_{3(3)} = -0.08 \quad \beta_8 = 0$
	$X_6 \sim N(20, 3)$	
	$X_7 \sim BIN(1, 0.5)$	$\beta_{4(1)} = 0.0015$
	$X_8 \sim N(0, 1)$	
$p = 15$	$X_1 \sim BIN(1, 0.4)$	$\beta_1 = 0.01 \quad \beta_{8(2)} = -0.06$
	$X_2 \sim BIN(1, 0.3)$	$\beta_2 = -0.15 \quad \beta_{9(1)} = 0.02$

**Tabel 3.3** Skenario untuk Analisis Data Simulasi (Lanjutan)

Jumlah variabel	Distribusi	Nilai Parameter	
$p = 15$	$X_3 \sim \text{BIN}(1, 0.5)$	$\beta_3 = -0.12$ $\beta_{9(2)} = -0.065$	
	$X_4 \sim \text{BIN}(1, 0.6)$	$\beta_4 = 0.012$ $\beta_{9(3)} = 0.01$	
	$X_5 \sim \text{BIN}(1, 0.7)$	$\beta_5 = 0.008$ $\beta_{10(1)} = -0.06$	
	$X_6 \sim \text{Mult}(4, 0.5, 0.3, 0.1, 0.1)$	$\beta_{6(1)} = -0.07$ $\beta_{10(2)} = 0.03$	
	$X_7 \sim \text{Mult}(4, 0.05, 0.3, 0.5, 0.05)$	$\beta_{6(2)} = 0,003$ $\beta_{11} = -0.04$	
	$X_8 \sim \text{Mult}(3, 0.3, 0.5, 0.2)$	$\beta_{6(3)} = -0,08$ $\beta_{12} = 0.015$	
	$X_9 \sim \text{Mult}(3, 0.4, 0.4, 0.2)$	$\beta_{7(1)} = 0.015$ $\beta_{13} = -0.015$	
	$X_{10} \sim \text{Mult}(4, 0.05, 0.3, 0.6, 0.05)$	$\beta_{7(2)} = 0.075$ $\beta_{14} = -0.01$	
	$X_{11} \sim N(40, 3)$	$\beta_{7(3)} = 0.07$ $\beta_{15} = -0.035$	
	$X_{12} \sim N(40, 2)$	$\beta_{8(1)} = 0.035$	
	$X_{13} \sim N(30, 2)$		
	$X_{14} \sim N(20, 2)$		
	$X_{15} \sim N(20, 3)$		
	$p = 17$	$X_1 \sim \text{BIN}(1, 0.4)$	$\beta_1 = 0.01$ $\beta_{8(2)} = -0.06$
		$X_2 \sim \text{BIN}(1, 0.3)$	$\beta_2 = -0.15$ $\beta_{9(1)} = 0.025$
$X_3 \sim \text{BIN}(1, 0.5)$		$\beta_3 = -0.12$ $\beta_{9(2)} = -0.065$	
$X_4 \sim \text{BIN}(1, 0.6)$		$\beta_4 = 0.012$ $\beta_{9(3)} = 0.01$	
$X_5 \sim \text{BIN}(1, 0.7)$		$\beta_5 = 0.008$ $\beta_{10(1)} = -0.06$	
$X_6 \sim \text{Mult}(4, 0.5, 0.3, 0.1, 0.1)$		$\beta_{6(1)} = -0.07$ $\beta_{10(2)} = 0.03$	
$X_7 \sim \text{Mult}(4, 0.05, 0.3, 0.5, 0.05)$		$\beta_{6(2)} = 0,003$ $\beta_{11} = -0.04$	
$X_8 \sim \text{Mult}(3, 0.3, 0.5, 0.2)$		$\beta_{6(3)} = -0,08$ $\beta_{12} = 0.015$	
$X_9 \sim \text{Mult}(3, 0.4, 0.4, 0.2)$		$\beta_{7(1)} = 0.015$ $\beta_{13} = -0.015$	
$X_{10} \sim \text{Mult}(4, 0.05, 0.3, 0.6, 0.05)$		$\beta_{7(2)} = 0.075$ $\beta_{14} = -0.01$	
$X_{11} \sim N(40, 3)$		$\beta_{7(3)} = 0.07$ $\beta_{15} = -0.035$	
$X_{12} \sim N(40, 2)$		$\beta_{8(1)} = 0.035$ $\beta_{16} = 0$	
$X_{13} \sim N(30, 2)$		$\beta_{17} = 0$	
$X_{14} \sim N(20, 2)$			
$X_{15} \sim N(20, 3)$			
$X_{16} \sim \text{BIN}(1, 0.5)$			
$X_{17} \sim N(0, 1)$			

Data riil yang digunakan yaitu data pasien kanker serviks yang pernah menjalani perawatan/pengobatan di RSUD Dr. Soetomo Surabaya. Data pasien kanker serviks diperoleh dari data rekam medik pasien kanker serviks. Kriteria inklusi dalam penelitian ini adalah sebagai berikut.

1. Pasien kanker serviks
2. Jenis kelamin : perempuan
3. Memiliki kelengkapan data pada rekam medik yang diperlukan dalam penelitian ini.

Adapun kriteria eksklusi dalam penelitian ini adalah pasien kanker serviks yang tidak memiliki kelengkapan data pada rekam medik. Sehingga dari 490 rekam medik, hanya 412 data yang lengkap (semua variabel yang digunakan terisi dalam rekam medik). Dimana dari 412 data tersebut ada 27 kasus pasien yang meninggal (tidak tersensor) dan 385 pasien lainnya tidak meninggal (tersensor). Variabel untuk data riil dalam penelitian ini adalah sebagai berikut.

**Tabel 3.4** Variabel Data Riil

Variabel	Keterangan	Definisi Operasional
$t$	<i>Survival time</i>	Waktu pertama kali pasien melakukan pengobatan/pemeriksaan di RSUD dr. Soetomo Surabaya hingga terjadinya <i>event</i> (meninggal) atau sampai waktu pengamatan berakhir jika tidak terjadi <i>event</i> (tidak meninggal).
Status	Status atau <i>event</i> 0 = tidak terjadi <i>event</i> 1= terjadi <i>event</i> (meninggal)	Pasien mengalami <i>event</i> (meninggal) atau tidak
$x_1$	Usia Pasien	Usia pasien waktu pertama kali pasien melakukan pengobatan.
$x_2$	Komplikasi 0 = Tidak Ada 1 = Ada	Ada atau tidaknya komplikasi penyakit lain selama melakukan pengobatan di RSUD dr. Soetomo Surabaya.
$x_3$	Status Anemia 0 = Tidak 1 = Ya	Anemia atau tidak selama melakukan pengobatan di RSUD dr. Soetomo.



**Tabel 3.5** Variabel Data Riil (Lanjutan)

Variabel	Keterangan	Definisi Operasional
$x_4$	Jenis Pengobatan 0 = Kemoterapi 1 = Transfusi PRC 2 = Keduanya 3 = Lainnya	Pengobatan adalah tindakan yang pernah dilakukan selama perawatan kanker serviks meliputi kemoterapi, transfusi PRC, keduanya, atau lainnya (selain kemoterapi dan transfusi PRC)
$x_5$	Stadium 1 = Stadium I 2 = Stadium II 3 = Stadium III 4 = Stadium IV	Stadium dimulai dari stadium 0 hingga stadium IV yang menyatakan informasi mengenai stadium klinis pasien kanker serviks pada saat pertama kali pasien datang.
$x_6$	Usia Saat Menikah	Usia pasien pada saat menikah
$x_7$	Usia pertama kali menstruasi	Usia pasien pada saat pertama kali menstruasi
$x_8$	Siklus menstruasi	Siklus pasien pada saat menstruasi
$x_9$	Lama menstruasi	Lama pasien mengalami menstruasi selamaa satu periode menstruasi
$x_{10}$	Frekuensi Melahirkan (Paritas)	Frekuensi pasien melahirkan anak mulai pada saat menikah hingga terkena kanker serviks dan melakukan pengobatan di RSUD dr. Soetomo Surabaya.
$x_{11}$	Riwayat KB 0 = Tidak 1 = Suntik 2 = Pil 3 = IUD	Alat kontrasepsi yang pernah digunakan oleh pasien.
$x_{12}$	Pendidikan 0 = SD 1 = SMP 2 = SMA 3 = Sarjana	Pendidikan yang telah ditempuh pasien.

### 3.3.2 Struktur Data

Struktur data penelitian berisi *layout* data yang nantinya akan dianalisis. Dalam Tabel 3.6 terdapat kolom subjek, *survival time*, status, dan variabel prediktor yang meliputi usia pasien, stadium, penyakit penyerta, komplikasi, jenis pengobatan, usia saat menikah, frekuensi melahirkan, alat kontrasepsi yang digunakan dan status anemia. Berdasarkan Tabel 3.6, kolom subjek berisi pasien

kanker serviks yang berobat di rumah sakit dr. Soetomo Surabaya tahun 2013 sampai dengan 2016.

**Tabel 3.6** Struktur Data Analisis *Survival* dengan *Time Independent Covariate*

Subjek	<i>Survival</i>	Status	Variabel Prediktor ( $x_{ij}$ )				
	<i>Time (T)</i>	( $\delta_j$ )	$X_1$	$X_2$	$X_3$	...	$X_p$
1	$t_1$	$\delta_1$	$x_{11}$	$x_{21}$	$x_{31}$		$x_{p1}$
2	$t_2$	$\delta_2$	$x_{12}$	$x_{22}$	$x_{32}$		$x_{p2}$
3	$t_3$	$\delta_3$	$x_{13}$	$x_{23}$	$x_{33}$		$x_{p3}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
$n - 1$	$t_{n-1}$	$\delta_{n-1}$	$x_{1(n-1)}$	$x_{2(n-1)}$	$x_{3(n-1)}$	...	$x_{p(n-1)}$
$n$	$t_n$	$\delta_n$	$x_{1n}$	$x_{2n}$	$x_{3n}$	...	$x_{pn}$

Keterangan :

$t_j$  = *Survival time* untuk subjek ke- $j$

$\delta_j$  = Status untuk subjek ke- $j$

$x_{ij}$  = Variabel prediktor ke- $i$  untuk subjek ke- $j$

$i = 1, 2, \dots, p ; j = 1, 2, \dots, n$

### 3.3.3 Langkah Analisis

Pada penelitian ini ada beberapa langkah tujuan yang ingin dicapai sehingga perlu dilakukan tahapan analisis. Berikut adalah tahapan analisis dengan menggunakan metode Cox PHM dan A-SURLSSVM.

#### 1. Untuk data simulasi

- Membangkitkan data dengan berbagai skenario sesuai dengan subbab (3.3.1) dengan menggunakan persamaan (2.62) untuk data dengan pola linier dan persamaan (2.63) untuk data dengan pola nonlinier.

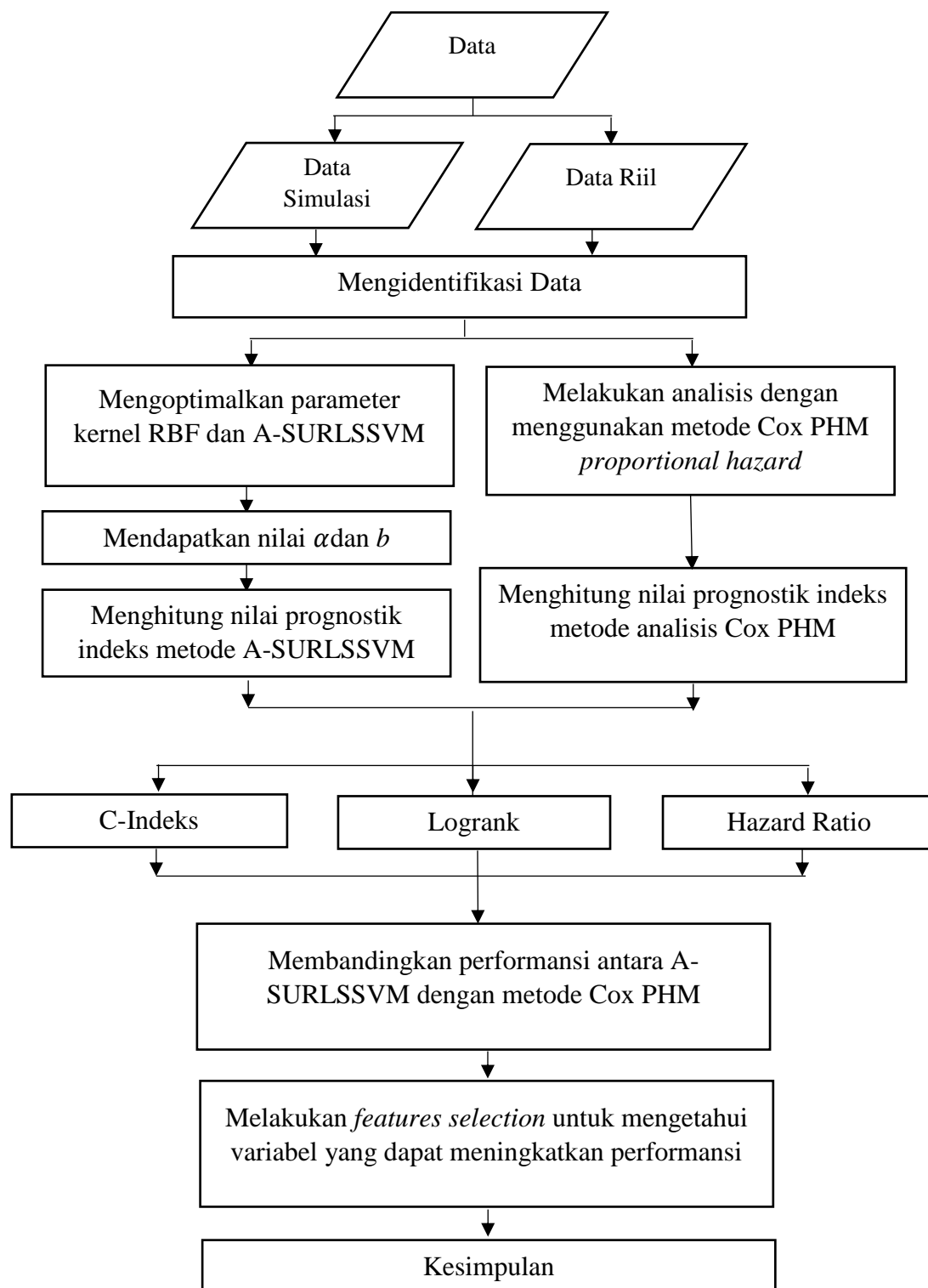
Untuk data riil

- Mengumpulkan dan mengidentifikasi data pasien kanker serviks di RSUD dr. Soetomo Surabaya tahun 2013 sampai dengan 2016 yang

akan digunakan untuk analisis sehingga terdapat 412 data pasien kanker serviks yang lengkap.

- Melakukan analisis statistika deskriptif pada 412 data pasien kanker serviks di RSUD dr. Soetomo.
2. Memasukkan data baik data simulasi maupun data riil pada algoritma subbab (3.1) sehingga dihasilkan prognostik indeks dari metode Cox PHM dan A-SURLSSVM dengan menggunakan persamaan (2.51).
  3. Menghitung nilai C-indeks, Logrank dan *Hazard Ratio* berdasarkan prognostik indeks untuk mengetahui performansi dari metode Cox PHM dan A-SURLSSVM.
  4. Membandingkan nilai C-indeks, Logrank dan *Hazard Ratio* antara metode metode A-SURLSSVM dengan metode Cox PHM.
  5. Melakukan *features selection* untuk mengetahui variabel yang dapat meningkatkan performansi dari model baik pada data riil maupun data simulasi.

Secara umum, Gambar 3.2 menunjukkan langkah-langkah analisis data.



**Gambar 3.2** Diagram Alir Langkah Analisis

*(halaman ini sengaja dikosongkan)*

## BAB 4

### ANALISIS DAN PEMBAHASAN

Pada bab ini diuraikan langkah-langkah dalam mencapai tujuan penelitian. Secara garis besar akan dijelaskan langkah-langkah dalam mendapatkan persamaan prognostik indeks dari model A-SURLSSVM, selanjutnya mengaplikasikan pada studi kasus pasien kanker serviks di RSUD dr. Soetomo Surabaya dan studi simulasi serta menggunakan *features selection* untuk mengetahui variabel yang berkontribusi meningkatkan performansi. Selanjutnya, membandingkan hasil model A-SURLSSVM dengan model Cox PHM.

#### 4.1 Kajian Teoritis Persamaan Prognostik indeks Model A-SURLSSVM

**Langkah pertama** yang dilakukan adalah mendapatkan nilai  $\alpha$  dan  $\beta$  yang optimal. Untuk mendapatkan nilai  $\alpha$  dan  $\beta$  yang optimal adalah dengan menyelesaikan persamaan kuadrat dengan membangun fungsi Lagrange dari fungsi tujuan dan fungsi kendala model A-SURLSSVM pada persamaan (2.47) dan (2.48).

$$\min_{\mathbf{w}, \xi, b, \chi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 + \frac{1}{2} \mu \sum_{j=1}^n \chi_j^2$$

$$\text{Fungsi kendala} \begin{cases} \mathbf{w}^T \boldsymbol{\varphi}(x_j) - \mathbf{w}^T \boldsymbol{\varphi}(x_i) = 1 + \xi_{ij}, \quad \forall i, j = 1, 2, \dots, n \\ \delta_j t_j = \delta_j (\mathbf{w}^T \boldsymbol{\varphi}(x_j) + b) - \chi_j, \quad \forall j = 1, 2, \dots, n \end{cases}$$

Dengan  $\gamma \geq 0$  dan  $\mu \geq 0$ . Berdasarkan fungsi tujuan dan fungsi kendala maka persamaan fungsi Langrange dari model A-SURLSSVM pada persamaan (2.49) adalah sebagai berikut:

$$L(\mathbf{w}, \xi, \chi, \alpha, \beta) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 + \frac{1}{2} \mu \sum_{j=1}^n \chi_j^2$$

$$- \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\mathbf{w}^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) - 1 - \xi_{ij})$$

$$- \sum_{j=1}^n \beta_j (\delta_j (-t_j + \mathbf{w}^T \boldsymbol{\varphi}(x_j) + b) - \chi_j)$$

**Langkah kedua** adalah dilakukan penurunan fungsi Lagrange terhadap  $\mathbf{w}, \xi, \chi, b, \alpha$  dan  $\beta$  pada persamaan (2.49) pada kondisi Karush Kuhn Tucker (KKT). Untuk menyelesaikan persamaan kuadrat, langkah pertama yang dilakukan adalah mengeliminasi  $\mathbf{w}, \xi$  dan  $\chi$  pada fungsi Lagrange dengan cara menurunkan persamaan fungsi Lagrange terhadap  $\mathbf{w}, \xi, \chi, b, \alpha$  dan  $\beta$ . Sehingga kondisi Karush Kuhn Tucker (KKT) untuk kondisi optimal digambarkan sebagai berikut:

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \\ \frac{\partial L}{\partial \xi_{ij}} = 0 \rightarrow \alpha_{ij} = -\gamma v_{ij} \xi_{ij} \\ \frac{\partial L}{\partial \chi_j} = 0 \rightarrow \beta_j = -\mu \chi_j \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{j=1}^n \beta_j \delta_j = 0 \\ \frac{\partial L}{\partial \alpha_{ij}} = 0 \rightarrow \mathbf{w}^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) - 1 - \xi_{ij} = 0 \\ \frac{\partial L}{\partial \beta_j} = 0 \rightarrow \delta_j (-t_j + \mathbf{w}^T \boldsymbol{\varphi}(x_j) + b) - \chi_j = 0 \end{array} \right.$$

**Langkah ketiga** adalah mengeliminasi  $\mathbf{w}, \xi$ , dan  $\chi$  dengan cara mensubstitusi hasil turunan fungsi Lagrange sehingga menjadi 3 persamaan linier pada persamaan (2.50).

$$\mathbf{w} = \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \quad (4.1)$$

$$\alpha_{ij} = -\gamma v_{ij} \xi_{ij} \text{ sehingga } \xi_{ij} = -\frac{\alpha_{ij}}{\gamma v_{ij}} \quad (4.2)$$

$$\beta_j = -\mu \chi_j \text{ sehingga } \chi_j = -\frac{\beta_j}{\mu} \quad (4.3)$$

$$\sum_{j=1}^n \beta_j \delta_j = 0 \quad (4.4)$$

$$\mathbf{w}^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) - 1 - \xi_{ij} = 0 \quad (4.5)$$

$$\delta_j (-t_j + \mathbf{w}^T \boldsymbol{\varphi}(x_j) + b) - \chi_j = 0 \quad (4.6)$$

mensubstitusikan  $\mathbf{w}$  pada persamaan (4.5) dan (4.6) dengan persamaan (4.1),  $\xi_{ij}$  pada persamaan (4.5) dengan persamaan (4.2) serta  $\chi_j$  pada persamaan (4.6) dengan persamaan (4.3). Sehingga didapatkan persamaan baru sebagai berikut:

$$\begin{aligned}
& \mathbf{w}^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) - 1 - \xi_{ij} = 0 \\
& \Leftrightarrow \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) - 1 - \left( -\frac{\alpha_{ij}}{\gamma v_{ij}} \right) = 0 \\
& \Leftrightarrow \left( \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) \right)^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) \right) + \left( \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \left( \frac{\alpha_{ij}}{\gamma v_{ij}} \right) = 1 \\
& \Leftrightarrow \left( \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i))^T \alpha_{ij}^T \right) (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n (\boldsymbol{\varphi}(x_j))^T \delta_j^T \beta_j^T (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \left( \frac{\alpha_{ij}}{\gamma v_{ij}} \right) = 1 \\
& \Leftrightarrow \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{\varphi}(x_j))^T - \boldsymbol{\varphi}(x_i)^T \alpha_{ij}^T (\boldsymbol{\varphi}(x_j)) - \sum_{i=1}^n \sum_{j=1}^n (\boldsymbol{\varphi}(x_j))^T - \boldsymbol{\varphi}(x_i)^T \alpha_{ij}^T (\boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n (\boldsymbol{\varphi}(x_j))^T \delta_j^T \beta_j^T (\boldsymbol{\varphi}(x_j)) \\
& \quad - \sum_{j=1}^n (\boldsymbol{\varphi}(x_j))^T \delta_j^T \beta_j^T (\boldsymbol{\varphi}(x_i)) + \left( \frac{\alpha_{ij}}{\gamma v_{ij}} \right) = 1
\end{aligned}$$

$$\delta_j (-t_j + \mathbf{w}^T \boldsymbol{\varphi}(x_j) + b) - \chi_j = 0$$

$$\begin{aligned}
& \Leftrightarrow \delta_j (-t_j + \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T \boldsymbol{\varphi}(x_j) + b) - \left( -\frac{\beta_j}{\mu} \right) = 0 \\
& \Leftrightarrow -\delta_j t_j + \delta_j \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) + \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T \boldsymbol{\varphi}(x_j) + \delta_j b + \frac{\beta_j}{\mu} = 0 \\
& \Leftrightarrow \delta_j \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) \right)^T \boldsymbol{\varphi}(x_j) + \delta_j \left( \sum_{j=1}^n \beta_j \delta_j \boldsymbol{\varphi}(x_j) \right)^T \boldsymbol{\varphi}(x_j) + \frac{\beta_j}{\mu} + \delta_j b = \delta_j t_j \\
& \Leftrightarrow \delta_j \left( \sum_{i=1}^n \sum_{j=1}^n \alpha_{ij} (\boldsymbol{\varphi}(x_j) - \boldsymbol{\varphi}(x_i)) \right)^T \boldsymbol{\varphi}(x_j) + \beta_j^T \left( \sum_{j=1}^n \delta_j (\boldsymbol{\varphi}(x_j))^T \delta_j^T \boldsymbol{\varphi}(x_j) + \frac{1}{\mu} \right) + \delta_j b = \delta_j t_j
\end{aligned}$$

$$\sum_{j=1}^n \beta_j \delta_j = 0$$

Setelah mengeliminasi  $\mathbf{w}$ ,  $\xi$  dan  $\chi$ , menjadi tiga persamaan baru, ketiga persamaan baru tersebut dapat dibuat ke dalam bentuk matriks. Diketahui bahwa  $\boldsymbol{\alpha}$  dan  $\boldsymbol{\beta}$  adalah Lagrange Multiplier yang berhubungan dengan konstrain,



$t = [t_1, t_2, \dots, t_n]^T$ ,  $\delta = [\delta_1, \delta_2, \dots, \delta_n]^T$  dan  $R \in \mathbb{R}^{n \times n}$  adalah diagonal matriks  $R = \text{diag}(\delta)$ . Sehingga didapatkan persamaan matriks sebagai berikut:

$$\begin{bmatrix} \mathbf{D K D}^T + \mathbf{I}_{n_c} / \gamma & \mathbf{D K R} & 0 \\ \mathbf{R K D}^T & \mathbf{R K R} + \mathbf{I}_n / \mu & \delta \\ 0 & \delta^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{\beta} \\ b \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{R t} \\ 0 \end{bmatrix} \quad (4.7)$$

**Langkah keempat** adalah mengoptimalkan parameter kernel RBF (*Radial Basis Function*) yaitu  $\sigma$  serta parameter  $\gamma$  dan  $\mu$  dari A-SURLSSVM. **Langkah kelima** adalah mendapatkan nilai  $\hat{a}$  dan  $\hat{b}$  berdasarkan langkah ketiga dan keempat. Pengoptimalan parameter dilakukan dengan mencari kombinasi nilai dari ketiga parameter yang dapat menghasilkan performansi *hazard ratio*, Logrank, dan C-index yang paling tinggi.

**Langkah keenam** adalah menghitung nilai prognostik indeks dengan menggunakan formula pada persamaan (2.51). Maka untuk mendapatkan prognostik indeks digunakan persamaan berikut.

$$\begin{aligned} \hat{u}(x) &= \hat{\mathbf{w}}^T \boldsymbol{\phi}(x) + b \\ &= \left( \sum_{i=1}^n \sum_{j=1}^n \hat{\alpha}_{ij} (\boldsymbol{\phi}(x_j) - \boldsymbol{\phi}(x_i)) + \sum_{j=1}^n \hat{\beta}_j \delta_j \boldsymbol{\phi}(x_j) \right)^T \boldsymbol{\phi}(x) + \hat{b} \\ &= (\hat{\boldsymbol{\alpha}}^T \mathbf{D} + \hat{\boldsymbol{\beta}}^T \mathbf{R}) \mathbf{K}_n(\mathbf{x}) + \hat{b} \end{aligned}$$

Diketahui bahwa  $\sum_{j=1}^n \hat{\beta}_j \delta_j = 0$ , sehingga untuk mendapatkan nilai

prognostik indeks yang optimal hanya perlu mencari nilai  $\hat{a}$  dan  $\hat{b}$  yang optimal.

Berikut diilustrasikan contoh perhitungan dari data awal yang sesuai dengan struktur data dalam subbab (3.3.2) hingga mendapatkan nilai prognostik indeks.

**Data:**

Subjek ke-	Survival Time (T)	Status Tersensor	X <sub>1</sub>	X <sub>2</sub>
1	2	0	40	1
2	1	0	41	0
3	5	1	39	0
4	3	1	58	1
5	4	1	31	0

### Perhitungan Matriks D:

Matriks **D** merupakan matrik berukuran  $n_e \times n$ ,  $n_e$  adalah selisih  $n$  dengan *event*/kejadian yang pertama setelah *survival time* ( $T$ ) yang telah diurutkan. Contoh perhitungan manual matriks **D** adalah sebagai berikut:

$$T = \begin{bmatrix} 2 \\ 1 \\ 5 \\ 3 \\ 4 \end{bmatrix} \quad \delta = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Adapun langkah-langkah untuk mendapatkan matriks **D** adalah sebagai berikut:

- Mengurutkan  $Y$ , sehingga vektor  $Y$  dan  $\delta$  menjadi

$$Y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} \quad \delta = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

- Mencari data yang pertama kali *event* terjadi. Dalam contoh di atas adalah data ke-3. Berdasarkan langkah kedua, maka didapatkan ukuran dari matriks **D** yaitu  $[n_e, n]$ . Dimana  $n_e$  adalah banyak data dikurangi urutan data yang pertama kali terjadi *event*. Sehingga  $n_e = 5 - 3 = 2$ . Sedangkan  $n$  adalah banyaknya data yaitu 5. Maka ukuran dari matriks **D** pada contoh di atas adalah  $[2,5]$
- Setelah mendapatkan ukuran dari matriks **D**, mengisi matriks **D** dengan angka nol

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- Mengisi matriks **D** yang berisi nol dengan 1 dan -1 sehingga jumlah setiap baris dari matriks tetap nol. Tata cara meletakkan nilai 1 dan -1 berdasarkan contoh di atas adalah
  - Pada baris pertama, pada kolom ke-3 karena *event* pertama terjadi pada data ke-3 dan diisi 1 karena kejadiannya setelah  $Y=5$ , sehingga pada kolom 5 diberi nilai -1

- b. Pada baris kedua, pada kolom ke-3 karena *event* pertama terjadi pada data ke-3 dan diisi -1 karena kejadiannya sebelum  $Y=4$ , sehingga pada kolom 5 diberi nilai 1.
- e. Sehingga didapatkan matriks  $\mathbf{D}$  sebagai berikut:

$$\mathbf{D} = \begin{bmatrix} 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 \end{bmatrix}$$

**Perhitungan Matriks Kernel ( $\mathbf{K}$ ) :**

Matriks  $\mathbf{K}$  merupakan matrik berukuran  $n \times n$ , dimana dalam penelitian ini kernel yang digunakan adalah kernel Radial Basis Function (RBF) dan terdapat pada persamaan (2.32). Contoh perhitungan matriks  $\mathbf{K}$  adalah sebagai berikut:

$$X = \begin{bmatrix} 4 & 1 \\ 4 & 0 \\ 2 & 0 \\ 5 & 1 \\ 3 & 0 \end{bmatrix}$$

Dengan menggunakan kernel parameter ( $\sigma = 0,5$ ) maka didapatkan matrik  $\mathbf{K}$  sebagai berikut.

$$\begin{bmatrix} 1,000000 & 0,367879 & 0,006738 & 0,367879 & 0,135335 \\ 0,367879 & 1,000000 & 0,018316 & 0,135335 & 0,367879 \\ 0,006738 & 0,018316 & 1,000000 & 4,54e-05 & 0,367879 \\ 0,367879 & 0,135335 & 4,54e-05 & 1,000000 & 0,006738 \\ 0,135335 & 0,367879 & 0,367879 & 0,006738 & 1,000000 \end{bmatrix}$$

**Perhitungan  $\hat{a}$  dan  $\hat{b}$  :**

Dengan menggunakan  $\gamma = 100$  dan  $\mu = 0,001$  serta perhitungan matriks  $\mathbf{D}$  dan  $\mathbf{K}$  maka didapatkan nilai  $\hat{a}$  dan  $\hat{b}$  sebagai berikut:

$$\hat{a} = \begin{bmatrix} 0 \\ 0 \\ 6,620522e-19 \\ -4,63762e-19 \\ -1,98289e-19 \end{bmatrix} \quad \hat{b} = 3,011763$$

### **Perhitungan prognostik indeks ( $\hat{u}$ ):**

Setelah mendapatkan nilai  $\hat{a}$  dan  $\hat{b}$  maka dapat dilakukan perhitungan prognostik indeks dengan menggunakan persamaan (2.51).

$$\hat{u} = \begin{bmatrix} 3,011763 \\ 3,011763 \\ 3,011763 \\ 3,011763 \\ 3,011763 \end{bmatrix}$$

Nilai prognostik inilah yang digunakan untuk menentukan subjek masuk ke dalam kategori *high risk* atau *low risk* serta digunakan untuk menghitung performansi dari model (*hazard ratio*, Logrank, dan C-indeks).

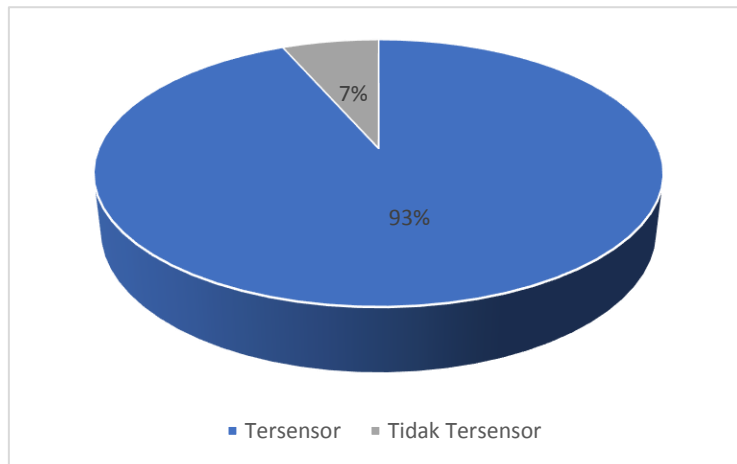
## **4.2 Penerapan Metode A-SURLSSVM dan Cox PHM pada Data Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya**

Analisis data pada penelitian ini diawali dengan menggambarkan karakteristik dari pasien kanker serviks dan faktor-faktor yang dianggap berpengaruh terhadap ketahanan pasien kanker serviks yang menjalanirawat inap di RSUD dr. Soetomo Surabaya, kurva survival Kaplan Meier dan uji Logrank, analisis survival dengan Cox PHM dan A-SURLSSVM serta membandingkan keduanya.

### **4.2.1 Karakteristik Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya**

Karakteristik pasien kanker serviks yang menjalani rawat jalan di RSUD dr. Soetomo Surabaya tersensor (tidak meninggal) dan tidak tersensor (meninggal) dapat dilihat pada Gambar 4.1.

Berdasarkan Gambar 4.1 dapat diketahui bahwa sebesar 7% atau sebanyak 27 pasien kanker serviks yang berobat rawat jalan di RSUD dr. Soetomo pada tahun 2013-2016 meninggal pada saat menjalani pengobatan. Sedangkan sebesar 93% atau sebanyak 385 pasien merupakan data tersensor (tidak meninggal) atau tidak mengalami *event*. Setiap pasien yang berobat memiliki *survival time* yang berbeda-beda.



Berikut ini adalah statistiak deskriptif *survival time* dari 412 pasien yang berobat di RSUD dr. Soetomo tahun 2013-2016.

**Tabel 4.1** Statistika Deskriptif dari *Survival Time*

Variabel	Rata-Rata	Median	Maksimum	Minimum	Standart Deviasi
T	121,56	94	712	3	96,48

Berdasarkan Tabel 4.1, dapat diketahui statistika deskriptif dari *survival time* 412 pasien kanker serviks yang berobat di RSUD dr. Soetomo Surabaya. Rata-rata pasien berobat selama 122 hari dengan kondisi akhir ada yang meninggal dunia dan ada yang tidak. Selisih pasien yang paling lama berobat (maksimum) dengan pasien yang paling cepat meninggalkan rumah sakit (minimum cukup besar). Hal ini juga terlihat dari standart deviasi *survival time* yang cukup besar yaitu 96,48.

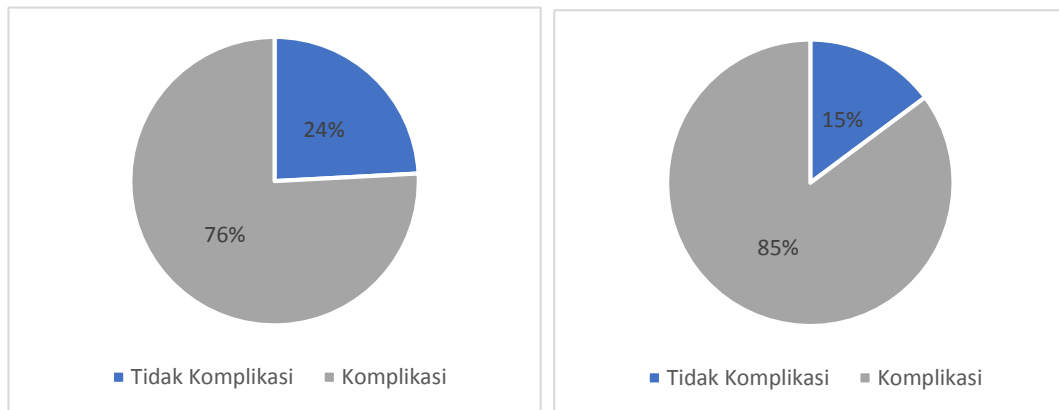
Berikut ini akan disajikan gambaran karakteristik pasien kanker serviks yang menjalani rawat jalan di RSUD dr. Soetomo Surabaya berdasarkan faktor-faktor yang diduga mempengaruhi kematian pasien kanker serviks. Analisis Statistika deskriptif untuk mengetahui karakteristik pasien kanker serviks berdasarkan faktor usia ( $X_1$ ), usia saat menikah ( $X_6$ ), usia pertama kali menstruasi ( $X_7$ ), siklus menstruasi ( $X_8$ ), lama menstruasi ( $X_9$ ), dan paritas ( $X_{10}$ ). Tabel 4.2 memberikan informasi rata-rata, nilai maksimum, nilai minimum, median, dan standar deviasi dari usia ( $X_1$ ), usia saat menikah ( $X_6$ ), usia pertama kali menstruasi ( $X_7$ ), siklus menstruasi ( $X_8$ ), lama menstruasi ( $X_9$ ), dan paritas ( $X_{10}$ ) antara pasien yang tersensor dan tidak tersensor.

**Tabel 4.2** Deskriptif Faktor  $X_1, X_6, X_7, X_8, X_9, X_{10}$

Deskriptif	$X_1$		$X_6$		$X_7$		$X_8$		$X_9$		$X_{10}$	
	0	1	0	1	0	1	0	1	0	1	0	1
Rata-Rata	49,15	50,04	19,00	20,10	13,29	13,10	28,40	28,67	6,49	6,22	2,72	2,81
Maksimum	80,00	69,00	34,00	30,00	18,00	17,00	35,00	30,00	12,00	7,00	8,00	6,00
Minimum	29,00	33,00	11,00	15,00	9,00	11,00	18,00	28,00	3,00	3,00	0,00	0,00
Median	49,00	51,00	19,00	19,00	13,00	13,00	28,00	28,00	7,00	7,00	3,00	3,00
Modus	49,00	51,00	20,00	20,00	12,00	12,00	28,00	28,00	7,00	7,00	2,00	3,00
Standar Deviasi	8,38	8,12	3,48	3,78	1,51	1,31	1,24	0,96	1,17	1,34	1,45	1,49

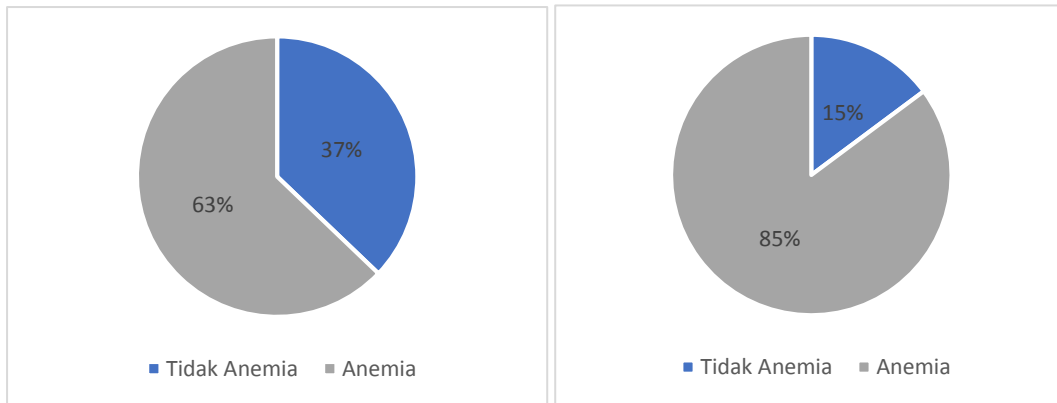
Pada Tabel 4.2 karakteristik pasien kanker serviks di RSUD dr. Soetomo Surabaya dari faktor usia ( $X_1$ ), usia saat menikah ( $X_6$ ), usia pertama kali menstruasi ( $X_7$ ), siklus menstruasi ( $X_8$ ), lama menstruasi ( $X_9$ ), dan paritas ( $X_{10}$ ) antara pasien yang tersensor dan tidak tersensor secara umum hampir sama dari segi rata-rata, median, modus, dan standar deviasi. Namun berbeda dari segi nilai maksimum dan minimumnya.

Analisis statistika deskriptif untuk mengetahui karakteristik pasien kanker serviks berdasarkan faktor komplikasi ( $X_2$ ), status anemia ( $X_3$ ), jenis pengobatan ( $X_4$ ), stadium ( $X_5$ ), riwayat KB ( $X_{11}$ ), dan pendidikan ( $X_{12}$ ) adalah sebagai berikut:



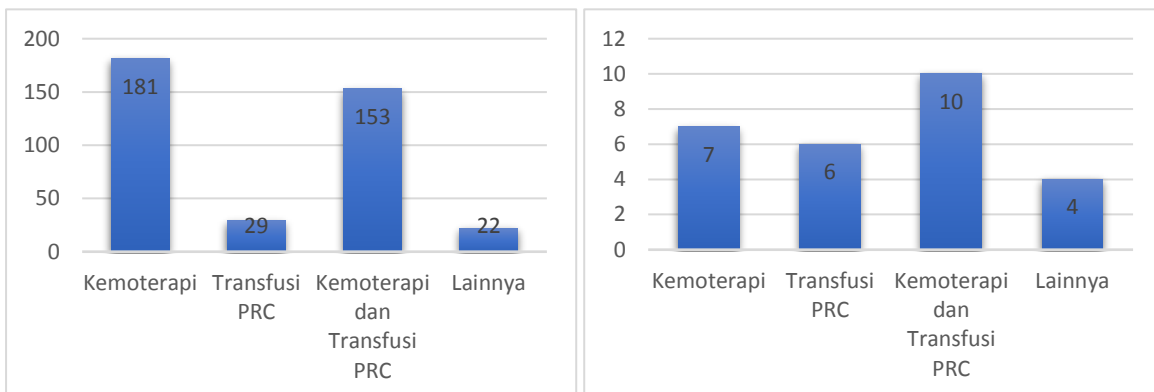
**Gambar 4.2** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Komplikasi

Gambar 4.2 menunjukkan prosentase pasien kanker serviks di RSUD dr. Soetomo yang tidak komplikasi dan yang komplikasi antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.2 dapat diketahui bahwa prosentase pasien yang memiliki komplikasi pada pasien yang meninggal lebih besar dari pada pasien yang tidak meninggal.



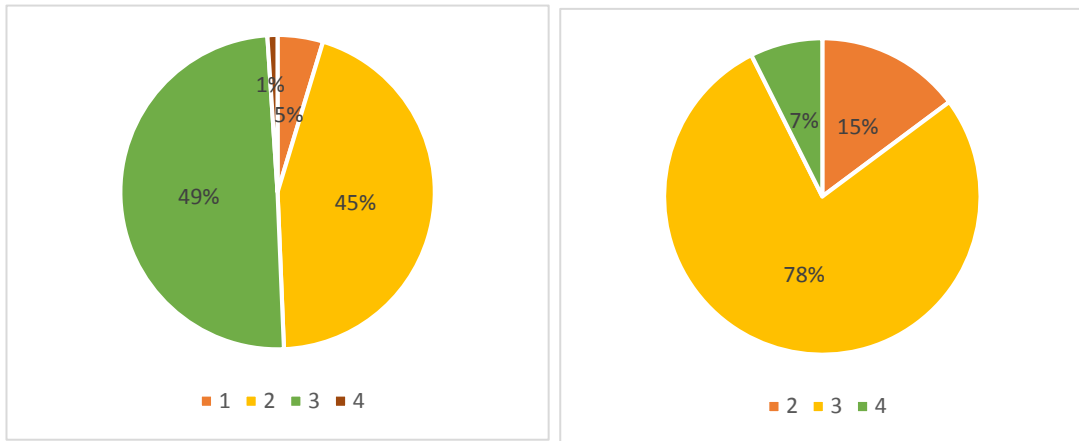
**Gambar 4.3** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Anemia

Gambar 4.3 menunjukkan prosentase pasien kanker serviks di RSUD dr. Soetomo yang tidak anemia dan yang anemia antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.3 dapat diketahui bahwa prosentase pasien yang memiliki anemia pada pasien yang meninggal lebih besar dari pada pasien yang tidak meninggal.



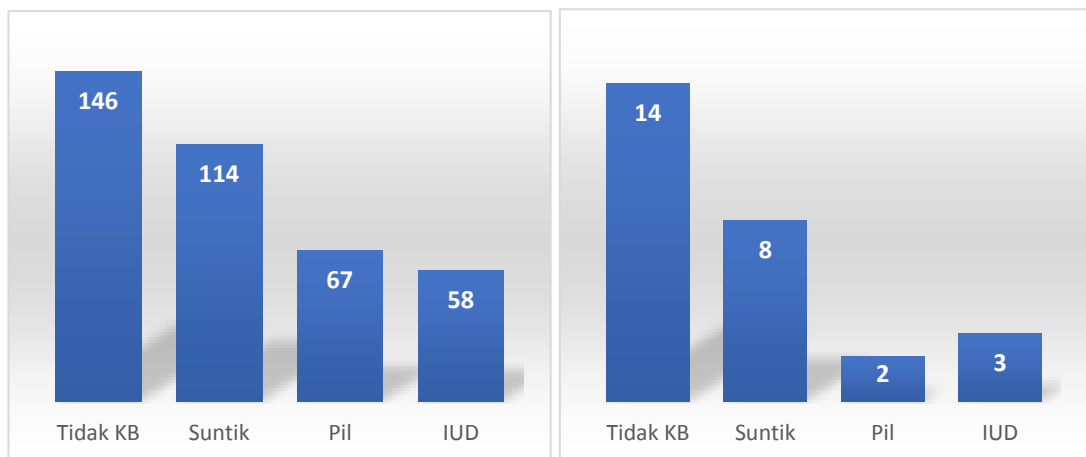
**Gambar 4.4** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Jenis Pengobatan

Gambar 4.4 menunjukkan jenis pengobatan pasien kanker serviks di RSUD dr. Soetomo antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.4 dapat diketahui bahwa pada pasien yang tidak meninggal, jenis pengobatan yang paling banyak dilakukan adalah kemoterapi. Sedangkan pada pasien yang meninggal pada saat pengobatan, jenis pengobatan yang paling banyak dilakukan adalah kemoterapi dan transfusi PRC (keduanya).



**Gambar 4.5** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Stadium

Gambar 4.5 menunjukkan prosentase stadium pasien kanker serviks di RSUD dr. Soetomo antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.5 dapat diketahui bahwa prosentase terbesar baik pada kelompok pasien yang meninggal dan tidak meninggal adalah stadium 3 dan terbesar kedua adalah stadium 2. Pada kelompok pasien yang tidak meninggal terdapat pasien yang masih masuk ke dalam stadium 1, sedangkan pada kelompok pasien yang meninggal hanya terdiri dari stadium 2, 3, dan 4.



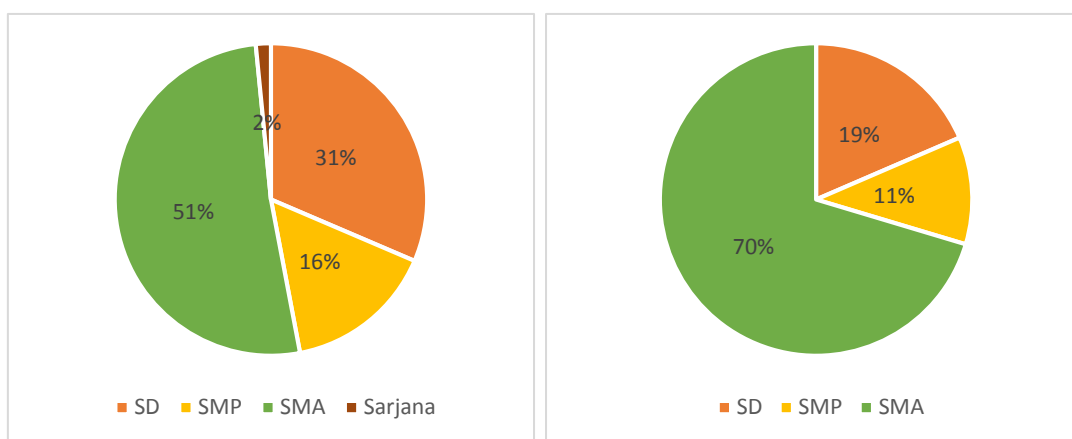
**Gambar 4.6** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Riwayat KB

Gambar 4.6 menunjukkan riwayat KB pasien kanker serviks di RSUD dr. Soetomo antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.6 dapat diketahui sebagian pasien tidak memiliki riwayat KB baik pada pasien yang masih hidup maupun yang



meninggal. Sedangkan KB yang paling banyak dijalani pasien pada kedua kelompok (tersensor dan tidak tersensor) adalah suntik.

Gambar 4.7 menunjukkan prosentase pendidikan pasien kanker serviks di RSUD dr. Soetomo antara pasien yang tidak meninggal (tersensor) dengan pasien yang meninggal (tidak tersensor). Berdasarkan Gambar 4.7 dapat diketahui bahwa prosentase pendidikan terbesar baik pada kelompok pasien yang meninggal dan tidak meninggal adalah SMA dan terbesar kedua adalah SD. Pada kelompok pasien yang tidak meninggal terdapat pasien yang memiliki riwayat pendidikan sarjana, sedangkan pada kelompok pasien yang meninggal hanya terdiri dari SD, SMP, dan SMA.



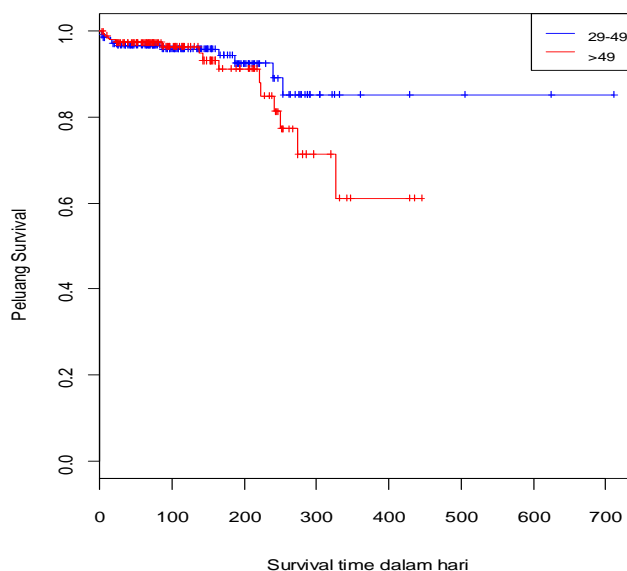
**Gambar 4.7** Karakteristik Pasien kanker Serviks yang Tersensor dan Tidak Tersensor dari Variabel Pendidikan

#### 4.2.2 Kurva Survival Kaplan Meier dan Uji Logrank

Kurva *survival* Kaplan Meier digunakan untuk mengetahui karakteristik kurva *survival* pasien kanker serviks berdasarkan faktor-faktor yang diduga mempengaruhi kematian pasien kanker serviks. Faktor tersebut diantaranya usia ( $X_1$ ), komplikasi ( $X_2$ ), status anemia ( $X_3$ ), jenis pengobatan ( $X_4$ ), stadium ( $X_5$ ), usia saat menikah ( $X_6$ ), usia pertama kali menstruasi ( $X_7$ ), siklus menstruasi ( $X_8$ ), lama menstruasi ( $X_9$ ), paritas ( $X_{10}$ ), riwayat KB ( $X_{11}$ ), dan pendidikan ( $X_{12}$ ). Sedangkan Logrank digunakan untuk membandingkan kurva *survival* dalam grup yang berbeda.

### Kurva Kaplan Meier dan Uji Logrank Usia ( $X_1$ )

Dari kurva Kaplan-Meier *survival function* faktor usia pada Gambar 4.8, garis biru menunjukkan kurva pasien kanker serviks yang berusia 29 hingga 49 tahun dan garis merah menunjukkan kurva pasien kanker serviks yang berusia di atas 49 tahun. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor usia saling berhimpit pada T 0-200 hari dan turun melambat serta kejadian pada saat T>200 hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang memiliki usia 29 hingga 49 tahun dengan pasien yang memiliki usia di atas 49 tahun adalah sama. Berikut ini disajikan kurva *survival* Kaplan Meier usia.



**Gambar 4.8** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Usia

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor usia.

**Tabel 4.3** Uji Logrank Faktor Usia

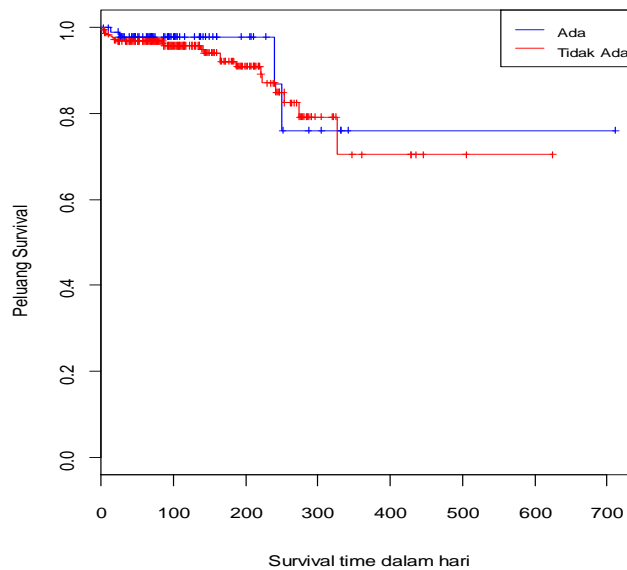
Logrank	df	P-Value
0.8	1	0,365

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.3, didapatkan nilai statistik uji 0,8 dengan derajat bebas 1 dan P-Value sebesar 0,365. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini

memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang memiliki usia 29 hingga 49 tahun dengan pasien yang memiliki usia di atas 49 tahun.

Kurva Kaplan Meier dan Uji Logrank Komplikasi ( $X_2$ )

Dari kurva Kaplan-Meier *survival function* faktor komplikasi pada Gambar 4.9, garis merah menunjukkan kurva pasien kanker serviks yang tidak komplikasi dan garis biru menunjukkan kurva pasien kanker serviks yang komplikasi. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor komplikasi saling berhimpit pada T 0-200 hari dan turun melambat serta kejadian pada saat T>200 hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang memiliki komplikasi maupun tidak adalah sama. Berikut ini disajikan kurva *survival* Kaplan Meier komplikasi.



**Gambar 4.9** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Komplikasi

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor komplikasi.

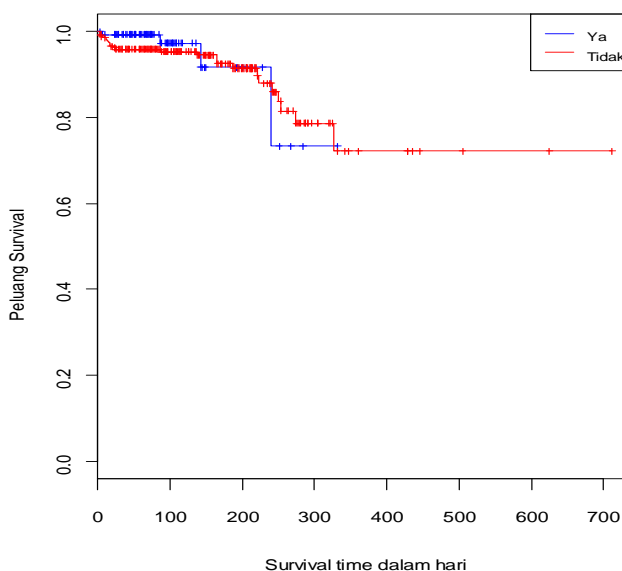
**Tabel 4.4** Uji Logrank Faktor Komplikasi

Logrank	df	P-Value
0,476	1	0,49

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.4, didapatkan nilai statistik uji 0,476 dengan derajat bebas 1 dan P-Value sebesar 0,49. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang memiliki komplikasi maupun tidak.

Kurva Kaplan Meier dan Uji Logrank Status Anemia ( $X_3$ )

Dari kurva Kaplan-Meier *survival function* faktor status anemia pada Gambar 4.10, garis merah menunjukkan kurva pasien kanker serviks yang tidak anemia dan garis biru menunjukkan kurva pasien kanker serviks yang anemia. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor anemia saling berhimpit pada T 0-200 hari dan turun melambat serta kejadian pada saat  $T > 200$  hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang anemia maupun tidak adalah sama. Berikut ini disajikan kurva *survival* Kaplan Meier status anemia.



**Gambar 4.10** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Status Anemia

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor anemia.

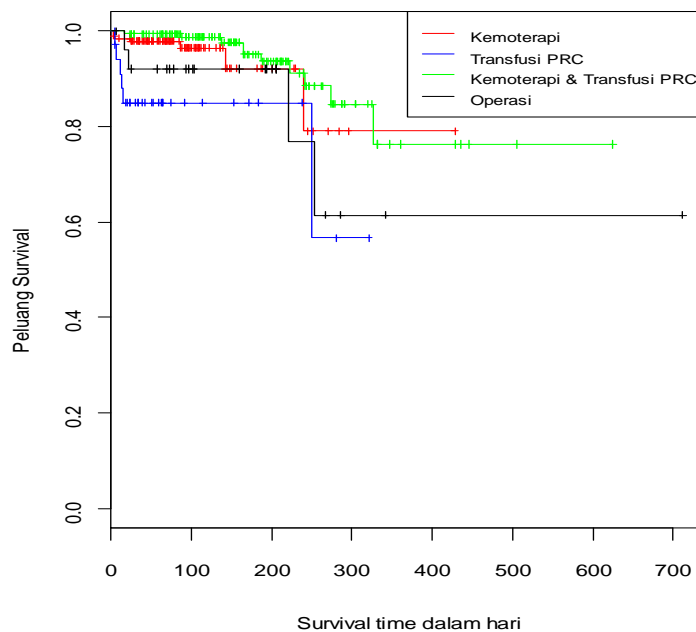
**Tabel 4.5** Uji Logrank Faktor Status Anemia

Logrank	Df	P-Value
0,897	1	0,344

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.5, didapatkan nilai statistik uji 0,897 dengan derajat bebas 1 dan P-Value sebesar 0,344. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang anemia maupun tidak.

Kurva Kaplan Meier dan Uji Logrank Jenis Pengobatan ( $X_4$ )

Berikut ini disajikan kurva survival Kaplan Meier jenis pengobatan.



**Gambar 4.11** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Jenis Pengobatan

Dari kurva Kaplan-Meier *survival function* faktor jenis pengobatan pada Gambar 4.11, garis merah menunjukkan kurva pasien kanker serviks yang jenis pengobatannya kemoterapi, garis biru menunjukkan kurva pasien kanker serviks yang jenis pengobatannya transfusi PRC, garis hijau menunjukkan kurva pasien kanker serviks yang jenis pengobatannya kemoterapi dan transfusi PRC sedangkan garis hitam menunjukkan kurva pasien kanker serviks yang jenis pengobatannya lainnya yaitu operasi. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor jenis pengobatan tidak saling berhimpit dan turun melambat. Artinya semakin lama melakukan perawatan probabilitas pasien kanker

serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* pasien yang jenis pengobatannya berbeda adalah tidak sama. Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor jenis pengobatan.

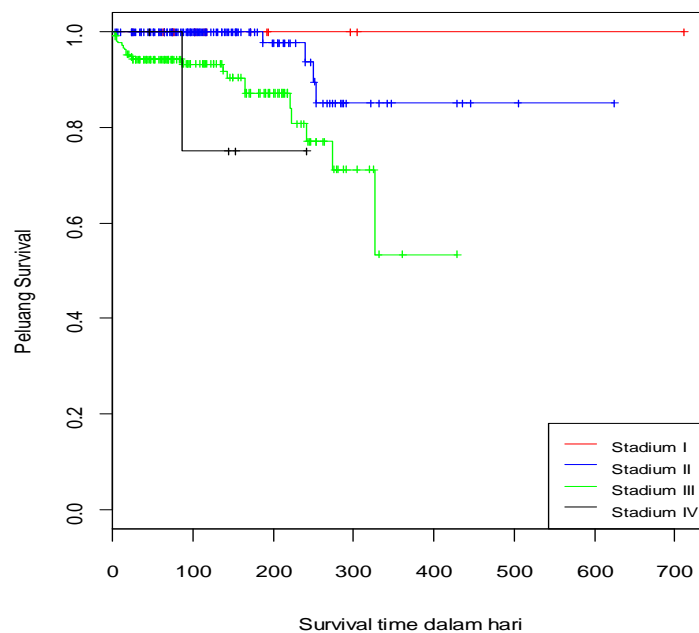
**Tabel 4.6** Uji Logrank Faktor Jenis Pengobatan

Logrank	df	P-Value
14,259	3	0,003

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.6, didapatkan nilai statistik uji 14,259 dengan derajat bebas 3 dan P-Value sebesar 0,003. Jika P-Value lebih kecil dibandingkan dengan nilai  $\alpha = 0,05$  maka tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa ada perbedaan kurva *survival* antar pasien berdasarkan jenis pengobatan.

Kurva Kaplan Meier dan Uji Logrank Stadium ( $X_5$ )

Berikut ini disajikan kurva survival Kaplan Meier stadium.



**Gambar 4.12** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Stadium

Dari kurva Kaplan-Meier *survival function* faktor stadium pada Gambar 4.12, garis merah menunjukkan kurva pasien kanker serviks dengan stadium 1, garis biru menunjukkan kurva pasien kanker serviks dengan stadium 2, garis hijau menunjukkan kurva pasien kanker serviks dengan stadium 3 sedangkan garis hitam

menunjukkan kurva pasien kanker serviks dengan stadium 4. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor stadium tidak saling berhimpit dan turun melambat. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* pasien dengan stadium berbeda adalah tidak sama. Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor stadium.

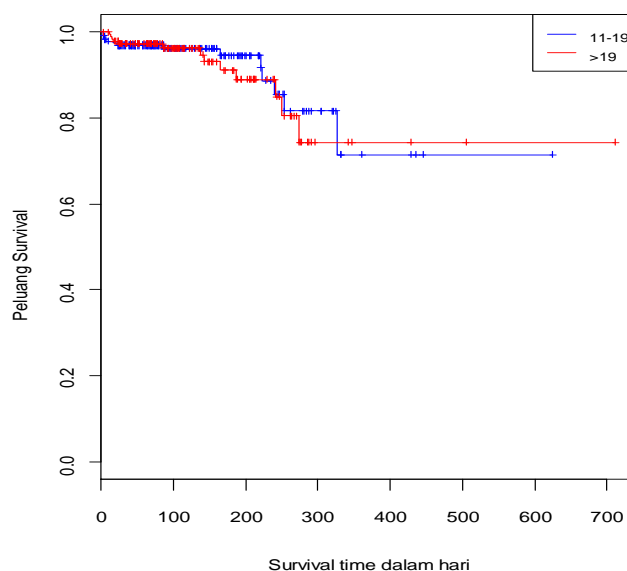
**Tabel 4.7** Uji Logrank Faktor Stadium

Logrank	Df	P-Value
15,199	3	0,002

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.7, didapatkan nilai statistik uji 15,199 dengan derajat bebas 3 dan P-Value sebesar 0,002. Jika P-Value lebih kecil dibandingkan dengan nilai  $\alpha = 0,05$  maka tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa ada perbedaan kurva *survival* antar pasien berdasarkan stadium.

Kurva Kaplan Meier dan Uji Logrank Usia saat menikah ( $X_6$ )

Berikut ini disajikan kurva survival Kaplan Meier usia saat menikah.



**Gambar 4.13** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Usia saat Menikah

Dari kurva Kaplan-Meier *survival function* faktor usia saat menikah pada Gambar 4.13, garis biru menunjukkan kurva pasien kanker serviks yang menikah pada usia 11 hingga 19 tahun dan garis merah menunjukkan kurva pasien kanker

serviks yang menikah di atas 19 tahun. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor usia saat menikah saling berhimpit pada T 0-300 hari dan turun melambat serta kejadian pada saat T>300 hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang menikah pada usia 11 hingga 19 tahun dengan pasien yang menikah di atas 19 tahun adalah sama. Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor usia saat menikah.

**Tabel 4.8** Uji Logrank Faktor Usia Saat Menikah

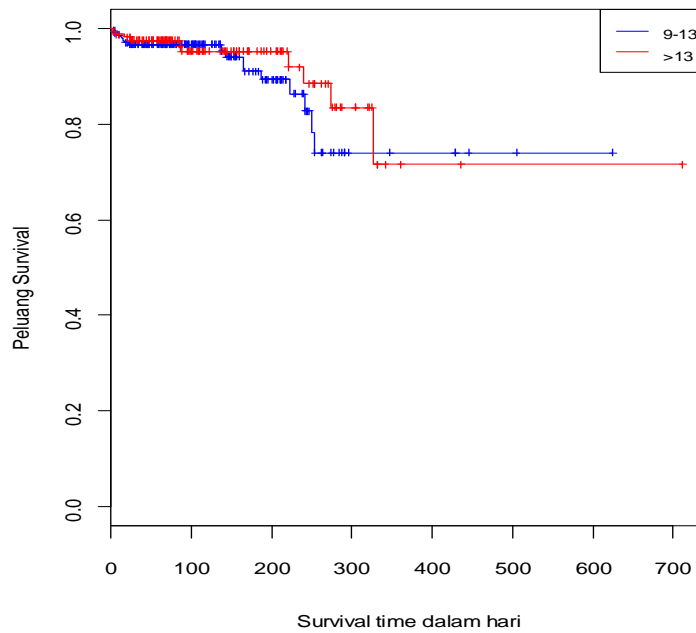
Logrank	Df	P-Value
0,2	1	0,689

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.8, didapatkan nilai statistik uji 0,2 dengan derajat bebas 1 dan P-Value sebesar 0,689. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang menikah pada usia 11 hingga 19 tahun dengan pasien yang menikah di atas 19 tahun.

*Kurva Kaplan Meier dan Uji Logrank Usia pertama kali menstruasi (X<sub>7</sub>)*

Dari kurva Kaplan-Meier *survival function* faktor usia pertama kali menstruasi pada Gambar 4.14, garis biru menunjukkan kurva pasien kanker serviks yang pertama kali menstruasi pada usia 9 hingga 11 tahun dan garis merah menunjukkan kurva pasien kanker serviks yang pertama kali menstruasi di atas 11 tahun. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor usia pertama kali menstruasi saling berhimpit pada T 0-200 hari dan turun melambat serta kejadian pada saat T>200 hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang pertama kali menstruasi pada usia 9 hingga 11 tahun dengan pasien yang menikah di atas 11 tahun adalah sama. Berikut ini disajikan kurva *survival* Kaplan Meier usia pertama kali menstruasi.





**Gambar 4.14** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Usia Pertama Kali Menstruasi

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor usia pertama kali menstruasi.

**Tabel 4.9** Uji Logrank Faktor Usia Pertama Kali Menstruasi

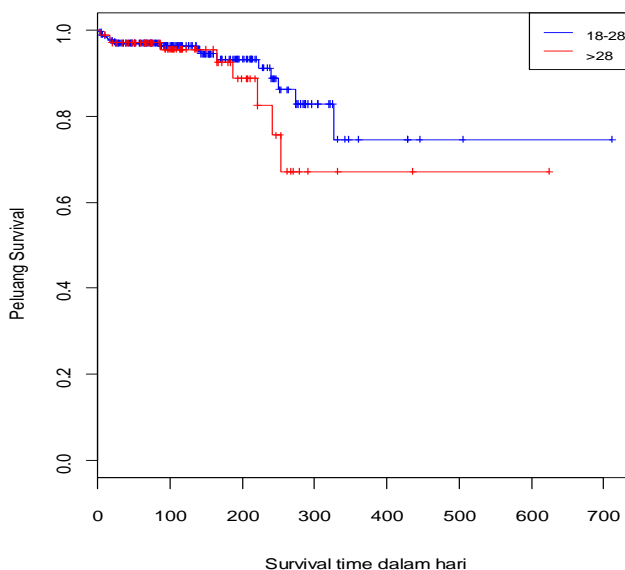
Logrank	Df	P-Value
0,7	1	0,415

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.9, didapatkan nilai statistik uji 0,7 dengan derajat bebas 1 dan P-Value sebesar 0,415. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang pertama kali menstruasi pada usia 9 hingga 11 tahun dengan pasien yang pertama kali menstruasi di atas 11 tahun.

Kurva Kaplan Meier dan Uji Logrank Siklus Menstruasi ( $X_8$ )

Dari kurva Kaplan-Meier *survival function* faktor siklus menstruasi pada Gambar 4.15, garis biru menunjukkan kurva pasien kanker serviks yang siklus menstruasinya 18-28 hari dan garis merah menunjukkan kurva pasien kanker serviks yang siklus menstruasinya di atas 28 hari. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor siklus menstruasi saling berhimpit

pada T 0-200 hari dan turun melambat serta kejadian pada saat T>200 hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang siklus menstruasinya 18 hingga 28 hari dengan pasien siklus menstruasinya di atas 28 hari adalah sama. Berikut ini disajikan kurva survival Kaplan Meier siklus menstruasi.



**Gambar 4.15** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Siklus Menstruasi

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor siklus menstruasi

**Tabel 4.10** Uji Logrank Faktor Siklus Menstruasi

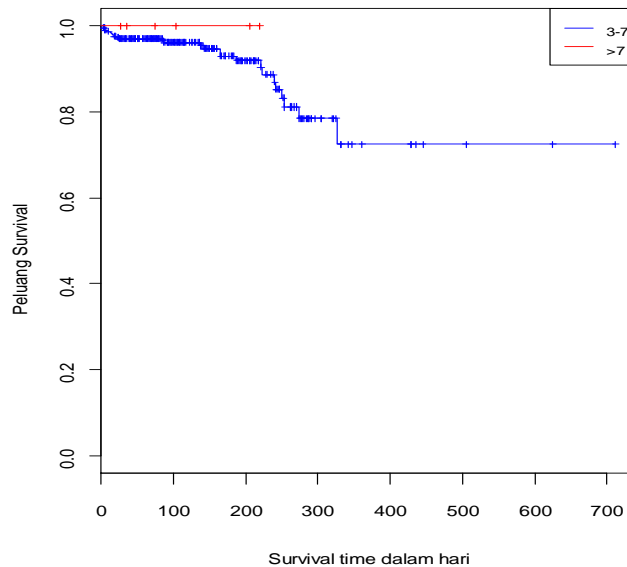
Logrank	df	P-Value
1	1	0,32

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.10, didapatkan nilai statistik uji 1 dengan derajat bebas 1 dan P-Value sebesar 0,32. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien siklus menstruasinya 18 hingga 28 hari dengan pasien yang siklus menstruasinya di atas 28 hari.

Kurva Kaplan Meier dan Uji Logrank Lama Menstruasi (X<sub>9</sub>)

Dari kurva Kaplan-Meier *survival function* faktor lama menstruasi pada Gambar 4.16, garis biru menunjukkan kurva pasien kanker serviks yang lama menstruasinya 3 hingga 7 hari dan garis merah menunjukkan kurva pasien kanker serviks yang lama menstruasinya di atas 7 hari. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor lama menstruasi tidak saling berhimpit, namun tidak ada *event* (pasien yang meninggal) yang lama menstruasinya lebih dari 7 hari. Artinya probabilitas untuk *survive* antara pasien yang lama menstruasinya 3 hingga 7 hari dengan pasien lama menstruasinya di atas 7 hari adalah sama.

Berikut ini disajikan kurva survival Kaplan Meier lama menstruasi.



**Gambar 4.16** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Lama Menstruasi

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor lama menstruasi.

**Tabel 4.11** Uji Logrank Faktor Lama Menstruasi

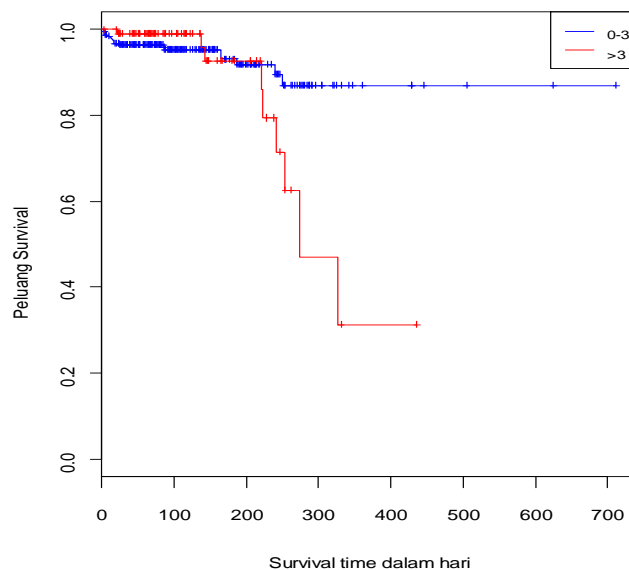
Logrank	Df	P-Value
0,3	1	0,565

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.11, didapatkan nilai statistik uji 0,3 dengan derajat bebas 1 dan P-Value sebesar 0,565. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini

memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang lama menstruasinya 3 hingga 7 hari dengan pasien yang lama menstruasinya di atas 7 hari.

#### Kurva Kaplan Meier dan Uji Logrank Paritas ( $X_{10}$ )

Dari kurva Kaplan-Meier *survival function* faktor paritas pada Gambar 4.17, garis biru menunjukkan kurva pasien kanker serviks yang frekuensi melahirkannya 0 hingga 3 anak dan garis merah menunjukkan kurva pasien kanker serviks yang anaknya lebih dari 3. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor paritas saling berhimpit pada T 0-250 hari dan turun melambat serta kejadian pada saat  $T > 250$  hari tidak terlalu banyak walaupun dua kurva tersebut tidak berhimpit. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* antara pasien yang paritasnya 0 hingga 3 anak dengan pasien yang anaknya lebih dari 3 adalah sama. Berikut ini disajikan kurva *survival* Kaplan Meier paritas.



**Gambar 4.17** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Paritas

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor paritas.

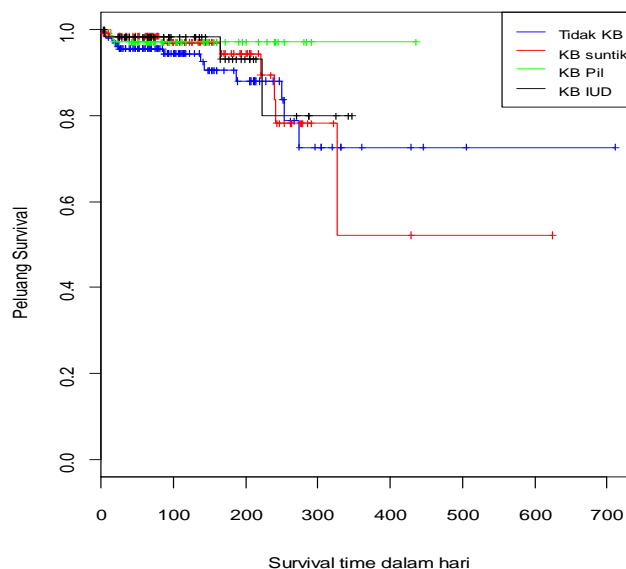
**Tabel 4.12** Uji Logrank Faktor Paritas

Logrank	Df	P-Value
1,7	1	0,197

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.12, didapatkan nilai statistik uji 1,7 dengan derajat bebas 1 dan P-Value sebesar 0,197. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antara pasien yang anaknya 0 hingga 3 dengan pasien yang anaknya lebih dari 3.

Kurva Kaplan Meier dan Uji Logrank Riwayat KB ( $X_{11}$ )

Berikut ini disajikan kurva survival Kaplan Meier riwayat KB.



**Gambar 4.18** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Riwayat KB

Dari kurva Kaplan-Meier *survival function* faktor riwayat KB pada Gambar 4.18, garis biru menunjukkan kurva pasien kanker serviks yang tidak KB, garis hijau menunjukkan kurva pasien kanker serviks yang KB suntik, garis krem menunjukkan kurva pasien kanker serviks yang KB pil sedangkan garis ungu menunjukkan kurva pasien kanker serviks yang KB IUD. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor riwayat KB tidak saling berhimpit dan turun melambat. Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* pasien dengan riwayat KB berbeda adalah sama.

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor riwayat KB.

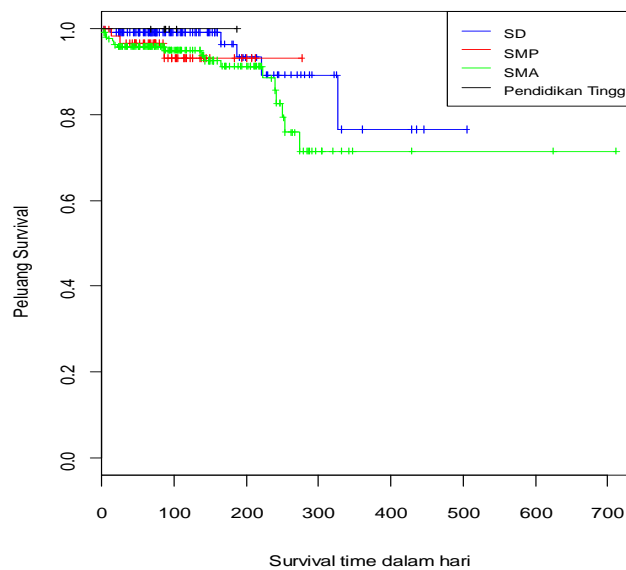
**Tabel 4.13** Uji Logrank Faktor Riwayat KB

Logrank	df	P-Value
2,374	3	0,498

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.13, didapatkan nilai statistik uji 2,374 dengan derajat bebas 3 dan P-Value sebesar 0,498. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antar pasien berdasarkan riwayat KB.

Kurva Kaplan Meier dan Uji Logrank Pendidikan ( $X_{12}$ )

Berikut ini disajikan kurva survival Kaplan Meier pendidikan.



**Gambar 4.19** Kurva *Survival* Kaplan Meier Pasien Kanker Serviks Pendidikan

Dari kurva Kaplan-Meier *survival function* faktor pendidikan pada Gambar 4.19, garis biru menunjukkan kurva pasien kanker serviks yang berpendidikan SD, garis hijau menunjukkan kurva pasien kanker serviks yang berpendidikan SMP, garis krem menunjukkan kurva pasien kanker serviks yang berpendidikan SMA sedangkan garis ungu menunjukkan kurva pasien kanker serviks yang berpendidikan sarjana. Dari kurva *survival* Kaplan Meier tersebut dapat dilihat bahwa kurva untuk faktor pendidikan tidak saling berhimpit dan turun melambat.

Artinya semakin lama melakukan perawatan probabilitas pasien kanker serviks untuk meninggal semakin rendah dan probabilitas untuk *survive* pasien dengan pendidikan berbeda adalah sama.

Untuk mendukung dugaan tersebut, maka perlu dilakukan uji Logrank untuk mengetahui ada tidaknya perbedaan yang berarti antara waktu *survival* pasien kanker serviks berdasarkan faktor pendidikan.

**Tabel 4.14** Uji Logrank Faktor Pendidikan

Logrank	df	P-Value
3,532	3	0,317

Berdasarkan hasil uji Logrank yang terdapat pada Tabel 4.14, didapatkan nilai statistik uji 3,532 dengan derajat bebas 3 dan P-Value sebesar 0,317. Jika P-Value lebih besar dibandingkan dengan nilai  $\alpha = 0,05$  maka gagal tolak  $H_0$ . Hal ini memberikan kesimpulan bahwa tidak ada perbedaan kurva *survival* antar pasien berdasarkan pendidikan.

Dari subbab 4.2.2 kurva kaplan Meier dan uji logrank didapatkan hasil bahwa variabel jenis pengobatan dan stadium adalah adanya perbedaan kurva *survival* pada variabel jenis pengobatan dan stadium. Artinya probabilitas pasien kanker serviks untuk *survive* pada masing-masing jenis pengobatan ataupun stadium tidak sama.

#### 4.2.3 Analisis Survival dengan Cox PHM dan A-SURLSSVM

Dalam penelitian ini, analisis survival dengan menggunakan Cox proportional hazard model (Cox PHM) digunakan untuk mengetahui faktor-faktor yang signifikan berpengaruh terhadap laju kematian pasien kanker serviks di RSUD dr. Soetomo Surabaya. Berikut hasil analisis yang diperoleh dengan menggunakan Cox PHM. Sebelum melakukan uji parsial, maka dilakukan uji serentak dengan menggunakan likelihood ratio test. Berdasarkan uji serentak tersebut, didapatkan nilai chi-square 50,5 berderajat bebas 20 dengan p-value 0,00019. Karena p-value lebih kecil dari  $\alpha=0,05$  maka tolak  $H_0$  yang artinya ada salah satu variabel yang signifikan mempengaruhi ketahanan pasien kanker serviks di RSUD dr. Soetomo Surabaya.

**Tabel 4.15** Cox PHM pada Data Kanker Serviks

Variabel	Koefisien	Standart error	Wald	P-Value
Usia (X <sub>1</sub> )	0,0135	0,0263	0,51	0,6085
Komplikasi (X <sub>2</sub> )	0,165	0,619	0,27	0,7893
Status Anemia (X <sub>3</sub> )	1,01	0,716	1,42	0,1563
Kemoterapi (X <sub>4(1)</sub> )	-0,462	0,766	-0,6	0,5466
Transfusi PRC (X <sub>4(2)</sub> )	0,217	0,776	0,28	0,7798
Keduanya (X <sub>4(3)</sub> )	-2,13	0,750	-2,85	0,0044
Stadium 1 (X <sub>5(1)</sub> )	-20,9	4800	0	0,9965
Stadium 2 (X <sub>5(2)</sub> )	-3,62	1,31	-2,75	0,006
Stadium 3 (X <sub>5(3)</sub> )	-1,06	1,09	-0,97	0,3306
Usia saat Menikah (X <sub>6</sub> )	0,0361	0,053	0,67	0,5016
Usia Pertama Kali Menstruasi (X <sub>7</sub> )	-0,134	0,157	-0,85	0,3937
Siklus Menstruasi (X <sub>8</sub> )	0,283	0,222	1,27	0,2026
Lama Menstruasi (X <sub>9</sub> )	-0,213	0,184	-1,16	0,2469
Paritas (X <sub>10</sub> )	-0,0779	0,155	-0,5	0,6141
Tidak KB (X <sub>11(1)</sub> )	0,585	0,711	0,82	0,4112
Suntik (X <sub>11(2)</sub> )	-0,0502	0,796	-0,06	0,9498
Pil (X <sub>11(3)</sub> )	-0,88	1,02	-0,86	0,3881
SD (X <sub>12(1)</sub> )	15,4	1630	0	0,9992
SMP (X <sub>12(2)</sub> )	16,5	1630	0	0,9992
SMA (X <sub>12(3)</sub> )	16,8	1630	0	0,9992

**Tabel 4.16** Asumsi Proportional Hazard Cox PHM pada Data Kanker Serviks

Variabel	Korelasi	P-Value	Keterangan
Usia (X <sub>1</sub> )	0,123557	0,539215	Tidak Melanggar
Komplikasi (X <sub>2</sub> )	-0,09039	0,653898	Tidak Melanggar
Status Anemia (X <sub>3</sub> )	-0,19291	0,335015	Tidak Melanggar
Kemoterapi (X <sub>4(1)</sub> )	0,008313	0,967174	Tidak Melanggar
Transfusi PRC (X <sub>4(2)</sub> )	-0,14021	0,485466	Tidak Melanggar
Keduanya (X <sub>4(3)</sub> )	0,071647	0,722491	Tidak Melanggar
Stadium 1 (X <sub>5(1)</sub> )	-0,854**	1,41E-08	Melanggar
Stadium 2 (X <sub>5(2)</sub> )	0,198476	0,320987	Tidak Melanggar
Stadium 3 (X <sub>5(3)</sub> )	-0,21223	0,2879	Tidak Melanggar
Usia saat Menikah (X <sub>6</sub> )	0,038923	0,847154	Tidak Melanggar
Usia Pertama Kali Menstruasi (X <sub>7</sub> )	0,001136	0,995511	Tidak Melanggar
Siklus Menstruasi (X <sub>8</sub> )	0,260994	0,188537	Tidak Melanggar
Lama Menstruasi (X <sub>9</sub> )	0,011295	0,955408	Tidak Melanggar
Paritas (X <sub>10</sub> )	0,642**	0,000307	Melanggar
Tidak KB (X <sub>11(1)</sub> )	0,017914	0,929333	Tidak Melanggar
Suntik (X <sub>11(2)</sub> )	0,156904	0,434463	Tidak Melanggar
Pil (X <sub>11(3)</sub> )	-0,2209	0,268184	Tidak Melanggar
SD (X <sub>12(1)</sub> )	0,109894	0,585296	Tidak Melanggar
SMP (X <sub>12(2)</sub> )	0,005711	0,977446	Tidak Melanggar
SMA (X <sub>12(3)</sub> )	-0,09512	0,636975	Tidak Melanggar



Berdasarkan Tabel 4.15, jenis pengobatan (kemoterapi dan transfusi PRC) dan stadium 2 signifikan berpengaruh terhadap laju kematian pasien kanker serviks di RSUD dr. Soetomo Surabaya. Sedangkan variabel yang lainnya tidak berpengaruh secara signifikan dalam mempengaruhi laju kematian pasien kanker serviks di RSUD dr. Soetomo Surabaya.

Kelemahan dari metode Cox PHM adalah asumsi proporsional hazard dan linearitas dalam kovariat. Berdasarkan Tabel 4.16 dapat diketahui bahwa ada dua variabel yaitu variabel stadium dan paritas tidak memenuhi (melanggar) asumsi proporsional hazard. Selanjutnya adalah memodelkan variabel jenis pengobatan dan stadium sebagai variabel yang signifikan.

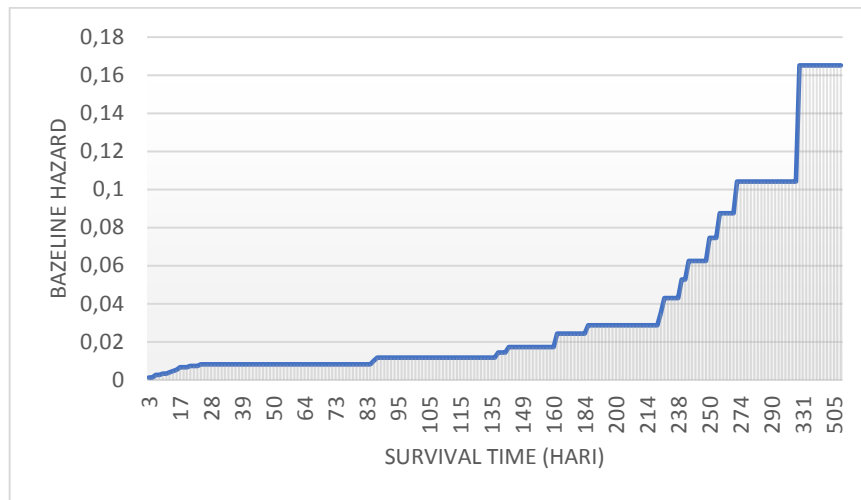
**Tabel 4.17** Cox PHM pada Data Kanker Serviks ( $X_4$  dan  $X_5$ )

Variabel	Koefisien	Standart error	Wald	P-Value	Keterangan
Kemoterapi ( $X_{4(1)}$ )	-1,20	0,65	-1,83	0,0669	Tidak Signifikan
Transfusi PRC ( $X_{4(2)}$ )	-0,26	0,67	-0,39	0,6974	Tidak Signifikan
Keduanya ( $X_{4(3)}$ )	-1,99	0,63	-3,15	0,0016	Signifikan
Stadium 1 ( $X_{5(1)}$ )	-19,30	5120	0,00	0,9970	Tidak Signifikan
Stadium 2 ( $X_{5(2)}$ )	-2,42	1,16	-2,08	0,0374	Signifikan
Stadium 3 ( $X_{5(3)}$ )	-0,41	1,04	-0,40	0,6902	Tidak Signifikan

Berdasarkan Tabel 4.17 dapat diketahui bahwa dengan menggunakan 2 variabel yang signifikan, dapat diketahui bahwa pada variabel jenis pengobatan (keduanya( $X_{4(3)}$ )) memiliki pvalue kurang dari 0,05 maka dapat dikatakan bahwa variabel tersebut berbeda dengan jenis pengobatan (lainnya) serta signifikan mempengaruhi laju kematian pasien kanker serviks. Untuk Stadium 2 ( $X_{5(2)}$ ) juga signifikan artinya variabel tersebut berbeda dengan stadium 4 dan juga secara signifikan mempengaruhi laju kematian pasien kanker serviks. Berdasarkan hasil parameter diperoleh model regresi Cox PHM sebagai berikut:

$$\hat{h}(t, X) = \hat{h}_0(t) \exp(-1,2x_{4(1)} - 0,262x_{4(2)} - 1,99x_{4(3)} - 19,3x_{5(1)} - 2,42x_{5(2)} - 0,413x_{5(3)})$$

Berdasarkan persamaan di atas didapatkan nilai bazeline hazard ( $\hat{h}_0(t)$ ) sebagai berikut:



**Gambar 4.20** Baseline Hazard dari Model Cox PHM ( $X_4$  dan  $X_5$ )

Berdasarkan Gambar 4.20 dapat diketahui baseline hazard pada saat ke  $t$  (dalam hari). Semakin lama seseorang pasien bertahan (*survival time*) maka kelajuan pasien tersebut untuk mengalami *event* semakin besar.

Karena dengan menggunakan seluruh variabel, ada pelanggaran asumsi *proportional hazard*, sehingga membutuhkan metode lainnya untuk menganalisis lebih lanjut. Selanjutnya adalah melakukan perbandingan performansi antara Cox PHM dengan A-SURLSSVM. Berikut adalah performansi Cox PHM dengan menggunakan tiga kriteria kebaikan yaitu Hazard Ratio, Logrank dan C-index.

**Tabel 4.18** Performansi Cox PHM pada Data Kanker Serviks

Hazard Ratio	Logrank	C-index
0,17	15,60	0,133

Dalam penelitian ini menggunakan metode dengan pendekatan SVM yaitu A-SURLSSVM digunakan untuk mengatasi kelemahan dari Cox PHM yaitu asumsi *proportional hazard*. Performansi A-SURLSSVM dengan berbagai kombinasi parameter  $C$ ,  $\mu$ , dan  $\gamma$  (metode *grid search*) sebanyak 125 kombinasi yang memaksimalkan ukuran performansi yang digunakan yaitu Hazard Ratio, Logrank, dan C-index yang didapatkan dari perhitungan prognostik indeks. Setelah didapatkan kombinasi parameter, maka dapat dilakukan perhitungan prognostik indeks. Untuk mendapatkan performansi yang paling paling bagus maka dipilih  $a$  dan  $b$  yang optimal.

$\alpha$  merupakan suatu vektor, sedangkan  $b$  merupakan suatu konstanta yang nilainya dapat dilihat pada lampiran 47. Untuk menghitung prognostik indeks individu ke- $i$  adalah

$$\hat{u}(\mathbf{x}_i) = \hat{\alpha}_i^T \mathbf{D} K(\mathbf{x}_i, \mathbf{x}) + 8,830366$$

Misalkan prognostik indeks untuk pasien ke-1 adalah

$$\hat{u}(\mathbf{x}_1) = -0,493601428 \mathbf{D} K(\mathbf{x}_1, \mathbf{x}) + 8,830366$$

Dengan  $\mathbf{x}_1$  merupakan kovariat dari pasien ke-1,  $\mathbf{x}_1 = [43, 0, 1, 0, \dots, 0, 0, 1]^T$  sehingga dari perhitungan tersebut didapatkan prognostik indeks sebesar 8,831710. Untuk keseluruhan prognostik indeks dapat dilihat pada lampiran 48. Statistika deskriptif prognostik indeks 412 pasien adalah sebagai berikut:

**Tabel 4.19** Statistika Deskriptif dari Prognostik Indeks

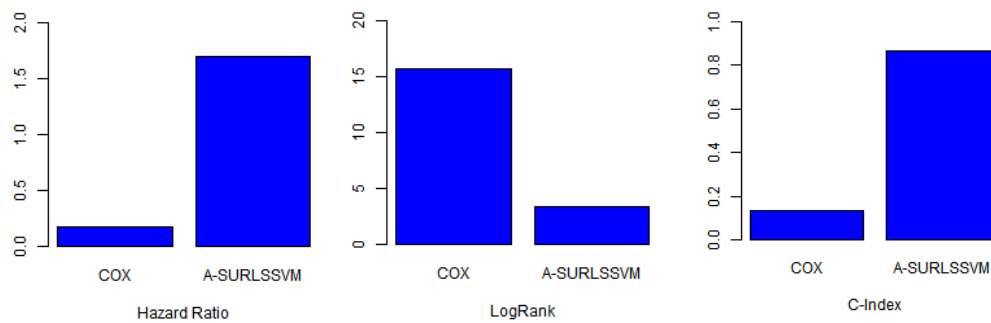
Variabel	Rata-Rata	Median	Standart Deviasi
Prognostik indeks	8,829	8,518	4,372

Prognostik indeks dari suatu pasien dapat dikategorikan ke dalam dua kategori yaitu pasien yang memiliki peluang yang besar untuk meninggal (*high risk*) dan pasien yang peluangnya kecil untuk mengalami *event*/ meninggal (*low risk*). Pengategorian didasarkan oleh nilai rata-rata prognostik indeks. Apabila prognostik indeks seorang pasien di atas rata-rata maka pasien tersebut masuk dalam kategori *high risk* dan sebaliknya. Berdasarkan Tabel 4.19, didapatkan rata-rata prognostik indeks sebesar 8,829. Jika nilai ini dibandingkan dengan prognostik indeks pasien yang pertama sebesar 8,831710, maka pasien pertama masuk ke dalam kategori *high risk*.

Setelah mendapatkan prognostik indeks dari 412 pasien, maka performansi dari A-SURLSSVM pada data kanker serviks dapat dihitung dan didapatkan hasil yang optimal adalah sebagai berikut:

**Tabel 4.20** Performansi A-SURLSSVM pada Data Kanker Serviks

Hazard Ratio	Logrank	C-index
1,689234	3,367676	0,8633166



**Gambar 4.21** Perbandingan Metode Cox PHM dan A-SURLSSVM pada Data Kanker Serviks

Berdasarkan hasil performansi pada Tabel 4.20 atau dapat dilihat pada Gambar 21 dapat diketahui bahwa dua dari tiga ukuran performansi yang digunakan yaitu Hazard Ratio dan C-index dari A-SURLSSVM lebih tinggi dari pada Cox PHM. Walaupun nilai Logrank Cox PHM lebih tinggi, nilai Logrank dari A-SURLSSVM juga dapat dikatakan sudah bagus. Perhitungan Logrank merupakan nilai chi-square berderajat bebas 1 karena nilai prognostik indeks dari Cox PHM maupun A-SURLSSVM dibagi ke dalam dua grup yaitu kelompok beresiko tinggi dan kelompok beresiko rendah. Dengan  $\alpha$  10% nilai dari chi square tabel berderajat bebas 1 adalah 2,706. Artinya jika nilai Logrank lebih dari 2,706 metode A-SURLSSVM yang digunakan, dipercaya 90% dapat membedakan grup yaitu grup beresiko rendah dan grup beresiko tinggi secara signifikan. Sehingga dapat disimpulkan bahwa dengan menggunakan data real (kanker serviks) metode A-SURLSSVM lebih bagus dari Cox PHM.

Dalam metode SVM tidak ada pengujian variabel yang dapat digunakan untuk mengetahui variabel-variabel yang berpengaruh terhadap respon, dalam kasus ini adalah laju kematian pasien kanker serviks di RSUD dr. Soetomo. Untuk mengetahui hal tersebut, dalam SVM dapat menggunakan seleksi variabel (*features selection*). Seleksi variabel yang digunakan dalam penelitian ini adalah *Backward elimination* yaitu dengan menghilangkan variabel satu per satu hingga mendapatkan performansi tertinggi. Berikut adalah hasil seleksi variabel dengan menggunakan metode *Backward elimination*.

**Tabel 4.21** Performansi A-SURLSSVM Hasil Seleksi Variabel

Tanpa	Hazard Ratio	Logrank	C-index dalam %
Usia ( $X_1$ )	1,707719	1,569268	83,01508
Komplikasi ( $X_2$ )	1,516527	0,956302	86,73367
Status Anemia ( $X_3$ )	1,492162	0,883024	86,0804
Jenis Pengobatan ( $X_4$ )	1,503784	0,916796	83,51759
Stadium ( $X_5$ )	1,516527	0,956302	83,41709
Usia saat Menikah ( $X_6$ )	1,461132	0,625297	78,59296
Usia Pertama Kali Menstruasi ( $X_7$ )	1,452930	0,771043	84,57286
Siklus Menstruasi ( $X_8$ )	1,516527	0,956302	79,69849
Lama Menstruasi ( $X_9$ )	1,527012	0,889924	84,92462
Paritas ( $X_{10}$ )	1,631624	1,322932	84,27136
Riwayat KB ( $X_{11}$ )	1,604338	1,232503	86,53266
Pendidikan ( $X_{12}$ )	1,529437	0,996871	85,87940

Tabel 4.21 menunjukkan hasil *features selection* dengan menghilangkan satu per satu variabel tidak dimasukkan dalam model. Dalam menentukan variabel mana yang perlu dihilangkan nilai yang digunakan adalah C-index. Dari 12 kali penghilangan variabel, variabel yang memiliki nilai C-index paling tinggi dan lebih besar dari 86.33166% maka variabel tersebut harus dihilangkan dalam model. Karena penghilangan variabel tersebut dapat meningkatkan nilai C-index dalam model. Dari dua belas variabel, penghilangan variabel komplikasi menghasilkan nilai C-index paling tinggi yaitu 86.73367%. Sehingga untuk *features selection* selanjutnya, variabel tersebut tidak dimasukkan lagi ke dalam model. Artinya variabel tersebut merupakan variabel yang paling tidak mempengaruhi waktu ketahanan hidup seseorang penderita kanker serviks di RSUD dr. Soetomo.

Setelah menghilangkan variabel komplikasi dalam model, dilakukan penghilangan satu per satu variabel untuk mengetahui performansi model jika tidak menggunakan variabel tersebut. Berdasarkan hasil analisis diketahui bahwa variabel riwayat KB memiliki nilai yang paling tinggi yaitu 87.73869%, dimana nilai tersebut lebih besar dari 86.73367%. Sehingga untuk analisis selanjutnya variabel riwayat KB tersebut tidak dimasukkan ke dalam model.

**Tabel 4.22** Performansi A-SURLSSVM Hasil Seleksi Variabel Setelah

Menghilangkan $X_2$			
Tanpa	Hazard Ratio	Logrank	C-index dalam %
Usia ( $X_1$ )	1,852978	2,064824	82,46231
<b>Komplikasi (<math>X_2</math>)</b>			
Status Anemia ( $X_3$ )	1,492162	0,883024	86,13065
Jenis Pengobatan ( $X_4$ )	1,418145	0,674064	85,67839
Stadium ( $X_5$ )	1,516527	0,956302	81,20603
Usia saat Menikah ( $X_6$ )	1,75816	1,748927	79,94975
Usia Pertama Kali Menstruasi ( $X_7$ )	1,354403	0,461569	85,07538
Siklus Menstruasi ( $X_8$ )	1,516527	0,956302	80,85427
Lama Menstruasi ( $X_9$ )	1,506787	0,925887	85,37688
Paritas ( $X_{10}$ )	1,631624	1,322932	84,42211
Riwayat KB ( $X_{11}$ )	1,492162	0,883024	87,73869
Pendidikan ( $X_{12}$ )	1,492162	0,883024	86,43216

Berdasarkan Tabel 4.23 dengan tidak melibatkan variabel komplikasi dan riwayat KB dapat diketahui tidak ada variabel yang memiliki C-index yang lebih besar dari 87.73869%, sehingga sepuluh variabel lainnya tetap masuk ke dalam model. Dengan demikian, dalam hal ini dengan menggunakan *features selection* dapat meningkatkan C-index sebesar 1,4%.

**Tabel 4.23** Performansi A-SURLSSVM Hasil Seleksi Variabel Setelah

Menghilangkan $X_2$ dan $X_{11}$			
Tanpa	Hazard Ratio	Logrank	C-index dalam %
Usia ( $X_1$ )	2,047877	3,237444	73,11558
<b>Komplikasi (<math>X_2</math>)</b>			
Status Anemia ( $X_3$ )	1,482648	0,854082	86,98492
Jenis Pengobatan ( $X_4$ )	1,387620	0,592022	85,67839
Stadium ( $X_5$ )	1,470327	0,817277	86,53266
Usia saat Menikah ( $X_6$ )	1,713205	1,54308	85,8794
Usia Pertama Kali Menstruasi ( $X_7$ )	1,38215	0,577162	85,42714
Siklus Menstruasi ( $X_8$ )	1,470327	0,817277	84,17085
Lama Menstruasi ( $X_9$ )	1,386153	0,530974	86,88442
Paritas ( $X_{10}$ )	1,550206	1,05671	85,22613
<b>Riwayat KB (<math>X_{11}</math>)</b>			
Pendidikan ( $X_{12}$ )	1,468477	0,813451	83,96985

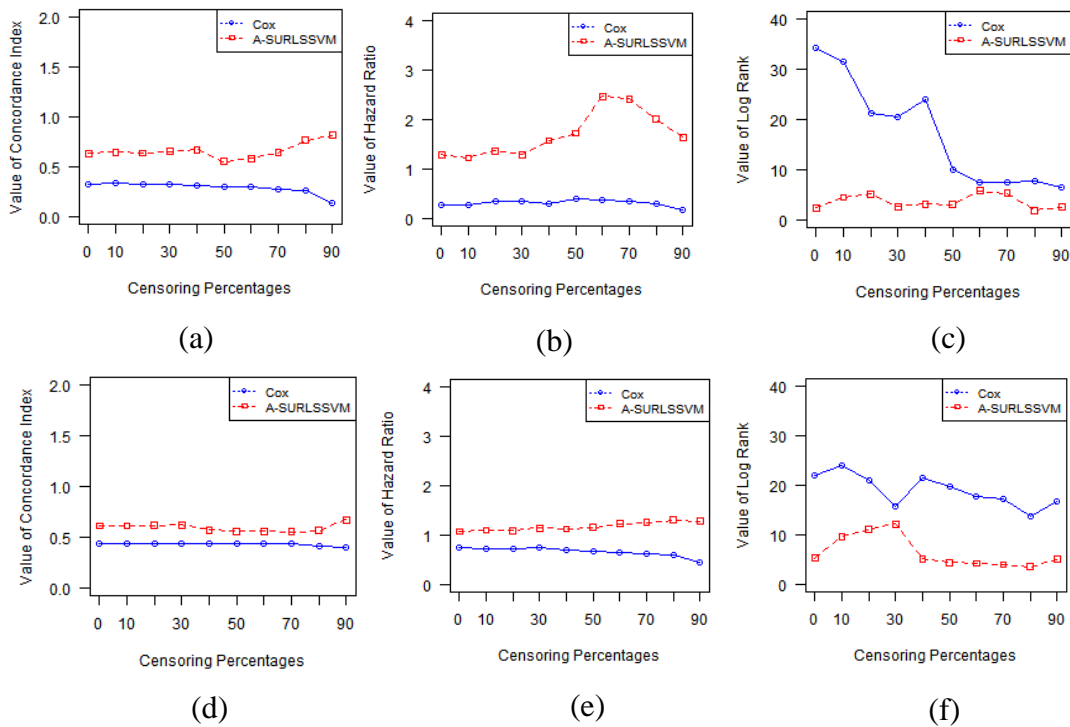
### 4.3 Penerapan Metode Cox PHM dan A-SURLSSVM pada Data Simulasi

Untuk mengetahui perbandingan performansi metode Cox PHM dan A-SURLSSVM dengan berbagai skenario/setting maka digunakan analisis data

dengan menggunakan data simulasi. Skenario yang digunakan dalam penelitian ini adalah data dibangkitkan dengan jumlah variabel prediktor yang berbeda-beda, data dibangkitkan dengan persentase censoring yang berbeda-beda, data dibangkitkan dengan jumlah data yang berbeda-beda, serta data *survival time* dibangkitkan dalam bentuk linier (tidak terdapat interaksi) dan nonlinier (terdapat interaksi). Berikut adalah hasil analisis dengan menggunakan data simulasi.

### 4.3.1 Penerapan dengan Persentase Sensoring dan Jumlah Data yang Berbeda pada Data dengan pola linier

Pada skenario ini *survival time* dibangkitkan dalam bentuk linier dengan persentase censoring 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 dan jumlah variabel sebanyak 6 serta jumlah data sebanyak 100 dan 1000. Selanjutnya data dianalisis menggunakan Cox PHM dan A-SURLSSVM. Performansi kedua metode dibandingkan dengan 3 ukuran yaitu C-index, Hazard Ratio, dan Logrank.



**Gambar 4.22** (a), (b), (c) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100  
 (d), (e), (f) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor 6 Jumlah Data 1000

Berdasarkan ketiga ukuran performansi yang digunakan pada Gambar 4.22 dapat diketahui bahwa dengan menggunakan concordance indek dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM pada  $n=100$  maupun pada  $n=1000$  karena nilainya lebih tinggi pada semua variasi persentase sensoring. Pola dari C-index pada Gambar 4.22 (a) dan (d) memiliki pola yang hampir sama. Sedangkan untuk pola dari Hazard Ratio pada Gambar 4.22 (b) dan (e) cukup berbeda. Itu artinya, C-index merupakan ukuran performansi yang cukup konsisten. Namun, berdasarkan nilai Logrank, metode Cox PHM memiliki Logrank yang lebih besar dari pada A-SURLSSVM. Nilai Logrank dari Cox PHM memiliki nilai yang besar pada data yang tidak terdapat data tersensor, dan memiliki pola menurun pada saat persentase sensoring dari data tersebut naik. Dari tiga ukuran performansi yang digunakan, ada dua ukuran yang menyimpulkan bahwa pada data dengan pola linier metode A-SURLSSVM lebih baik dari Cox. Hasil yang sama juga didapatkan dengan menggunakan variabel prediktor 8, 15, dan 17 yang dapat dilihat pada lampiran.

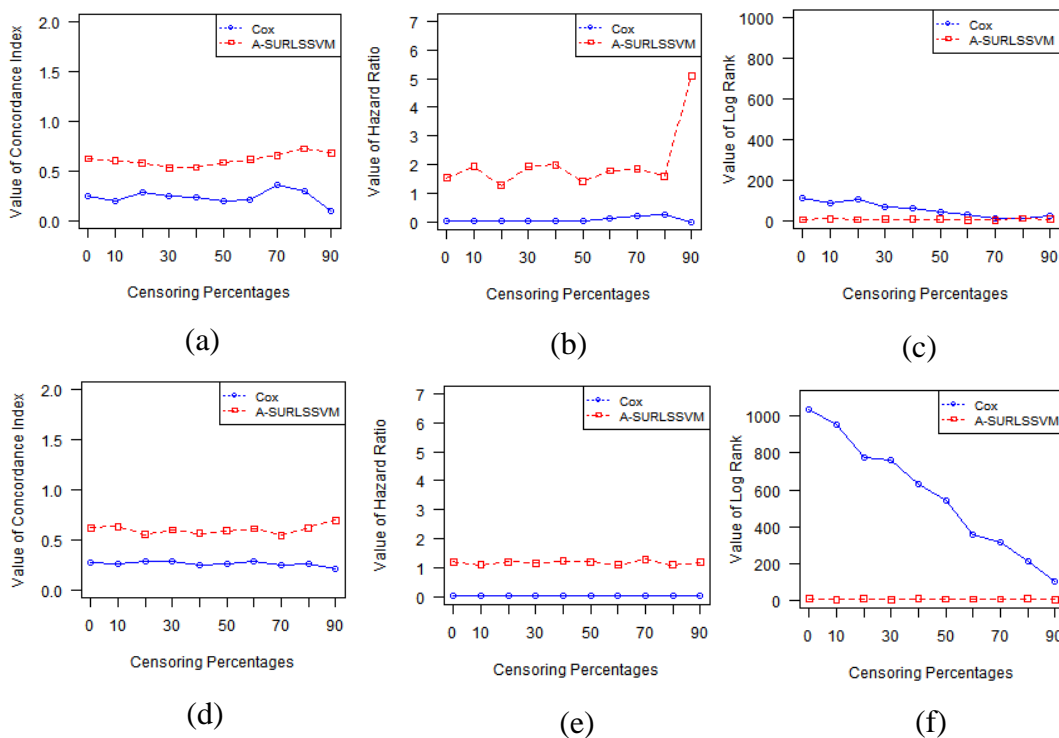
#### **4.3.2 Penerapan dengan Persentase Sensoring dan Jumlah Data yang Berbeda pada Data dengan pola nonlinier**

Pada skenario ini *survival time* dibangkitkan dalam bentuk nonlinier dengan persentase sensoring 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 dan jumlah variabel sebanyak 6 serta 100 dan 1000. Selanjutnya data dianalisis menggunakan Cox PHM dan A-SURLSSVM. Performansi kedua metode dibandingkan dengan 3 ukuran yaitu C-index, Hazard Ratio, dan Logrank.

Berdasarkan ketiga ukuran performansi yang digunakan dapat diketahui bahwa dengan menggunakan concordance indek dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM pada  $n=100$  maupun pada  $n=1000$  karena nilainya lebih tinggi pada semua variasi persentase sensoring. Pola dari C-index pada Gambar 4.23 (a) dan (d) memiliki pola yang hampir sama. Sedangkan untuk pola dari hazard rati pada Gambar 4.23 (b) dan (e) cukup berbeda. Itu artinya, Namun, berdasarkan nilai Logrank, metode Cox PHM memiliki Logrank yang lebih besar dari pada A-SURLSSVM. Nilai Logrank dari Cox PHM memiliki nilai yang besar pada data yang tidak terdapat data tersensor, dan memiliki pola menurun pada



saat persentase sensing dari data tersebut naik. Selain itu, semakin besar jumlah data yang digunakan, nilai Logrank dari Cox PHM juga semakin besar. Berbeda halnya dengan C-index yang nilainya cenderung konsisten walaupun jumlah datanya berubah. C-index merupakan ukuran performansi yang cukup konsisten. Dari tiga ukuran performansi yang digunakan, ada dua ukuran yang menyimpulkan bahwa pada data dengan pola nonlinier metode A-SURLSSVM lebih baik dari Cox. Hasil yang sama juga didapatkan dengan menggunakan variabel prediktor 8, 15, dan 17 yang dapat dilihat pada lampiran.

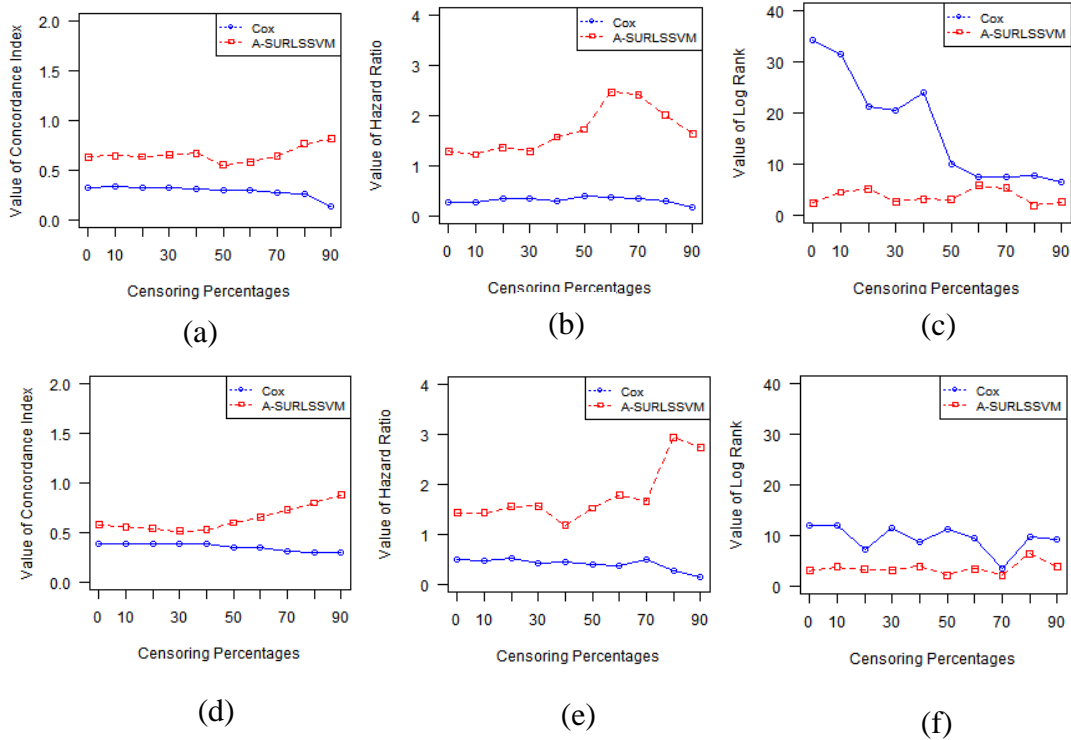


**Gambar 4.23** (a), (b), (c) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100  
 (d), (e), (f) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor 6 Jumlah Data 1000

### 4.3.3 Penerapan dengan Persentase Sensoring dan Jumlah Variabel yang Berbeda pada Data dengan pola linier

Pada skenario ini *survival time* dibangkitkan dalam bentuk linier dengan persentase sensing 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 dan jumlah variabel 6, dan

8 serta jumlah data sebanyak 100. Selanjutnya data dianalisis menggunakan Cox PHM dan A-SURLSSVM. Performansi kedua metode dibandingkan dengan 3 ukuran yaitu C-index, Hazard Ratio, dan Logrank.



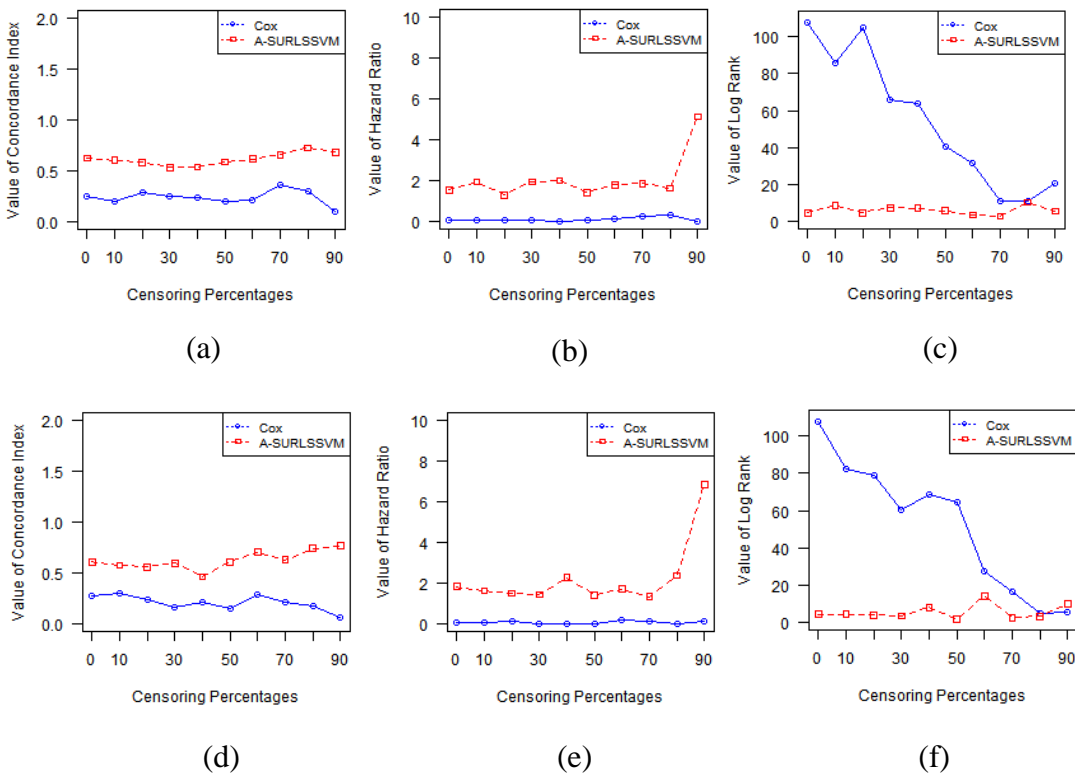
**Gambar 4.24** (a), (b), (c) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100  
 (d), (e), (f) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor 8 Jumlah Data 100

Berdasarkan ketiga ukuran performansi yang digunakan dapat diketahui bahwa dengan menggunakan C-index dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM karena nilainya lebih tinggi pada variabel 6 dan 8 serta 15 dan 17 yang dapat dilihat pada lampiran. Penambahan jumlah variabel prediktor dari 6 menjadi 8 dengan menyimulasikan  $\beta = 0$  pada  $X_7$  dan  $X_8$  membentuk pola C-index dan Hazard Ratio yang sama pada persentase censoring 0 sampai 90% yang dapat dilihat pada Gambar 4.23 (a)(b)(c)(d). Selain itu, dapat diketahui bahwa semakin besar persentase censoring, akan menghasilkan C-index pada metode A-SURLSSVM yang lebih tinggi atau lebih bagus terutama pada data yang persentase censoringnya di atas 50%. Dimana hal tersebut berbeda dengan performansi dari metode Cox PHM. Namun, berbeda halnya pada pola Logrank. Nilainya semakin turun saat persentase censoringnya semakin besar. Hasil yang sama

juga didapatkan pada simulasi dengan jumlah data 1000 dan 5000 yang dapat dilihat pada lampiran.

#### 4.3.4 Penerapan dengan Persentase Sensoring dan Jumlah Variabel yang Berbeda pada Data dengan pola nonlinier

Pada skenario ini *survival time* dibangkitkan dalam bentuk nonlinier dengan persentase sensing 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 dan jumlah variabel 6 dan 8 serta jumlah data sebanyak 100. Selanjutnya data dianalisis menggunakan Cox PHM dan A-SURLSSVM. Performansi kedua metode dibandingkan dengan 3 ukuran yaitu C-index, Hazard Ratio, dan Logrank.



**Gambar 4.25** (a), (b), (c) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor 6 Jumlah Data 100

(d), (e), (f) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor 8 Jumlah Data 100

Berdasarkan ketiga ukuran performansi yang digunakan dapat diketahui bahwa dengan menggunakan C-index dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM karena nilainya lebih tinggi pada variabel 6 dan 8 serta 15 dan 17 yang dapat dilihat pada lampiran. Penambahan jumlah variabel prediktor dari 6 menjadi 8 dengan menyimulasikan  $\beta = 0$  pada  $X_7$  dan  $X_8$

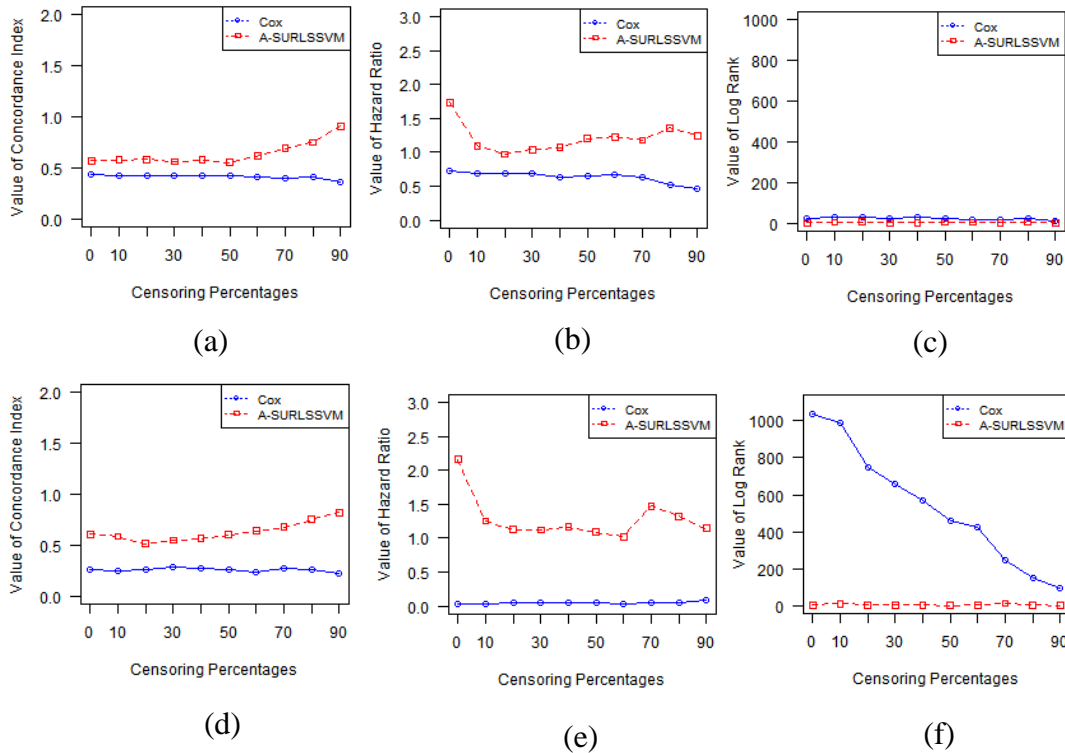
membentuk pola C-index, Hazard Ratio serta Logrank yang sama pada persentase censoring 0 sampai 90%. Namun, Logrank dari Cox PHM memiliki nilai yang besar pada data yang tidak terdapat data tersensor, dan memiliki pola menurun pada saat persentase censoring dari data tersebut naik. Dari tiga ukuran performansi yang digunakan, ada dua ukuran yang menyimpulkan bahwa pada data dengan pola nonlinier metode A-SURLSSVM lebih baik dari Cox. Hasil yang sama juga didapatkan pada simulasi dengan jumlah data 1000 dan 5000 yang dapat dilihat pada lampiran.

#### **4.3.5 Penerapan dengan Persentase Sensoring yang Berbeda pada Data dengan pola linier dan Nonlinier**

Pada skenario ini *survival time* dibangkitkan dalam bentuk linier dan nonlinier dengan persentase censoring 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 dan jumlah variabel 17 dan jumlah data 1000.

Berdasarkan ketiga ukuran performansi yang digunakan dapat diketahui bahwa dengan menggunakan C-index dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM karena nilainya lebih tinggi baik pada data dengan pola linier maupun nonlinier. Pola dari C-index dan Hazard Ratio pada data dengan pola linier dan nonlinier hampir sama. Dengan menggunakan A-SURLSSVM, pada data dengan pola linier, nilai C-index dan Hazard Ratio hampir sama dengan nonlinier. Namun, pada model Cox PHM nilai C-index dan Hazard Ratio selisihnya cukup besar antara data dengan pola linier dan nonlinier. C-index dan Hazard Ratio dari Cox PHM pada data dengan pola nonlinier lebih rendah dari pada data dengan pola linier. Sehingga pada data dengan pola nonlinier performansi dari A-SURLSSVM jauh lebih bagus dibandingkan Cox PHM. Sebaliknya, jika dilihat dari Logrank, nilainya Logrank cox pada data dengan pola nonlinier jauh lebih tinggi dibandingkan A-SURLSSVM. nilai Logrank dari metode A-SURLSSVM cenderung konstan untuk semua persentase censoring baik pada data dengan pola linier maupun data dengan pola nonlinier. Sedangkan pada Cox PHM, pada data yang persentase censoringnya kecil, nilai Logranknya sangat tinggi dan menurun saat persentase censoringnya naik dan terlihat sangat jelas pada data dengan pola

nonlinier. Hasil yang sama juga diperoleh pada jumlah data dan jumlah variabel lainnya yang dapat dilihat pada lampiran.



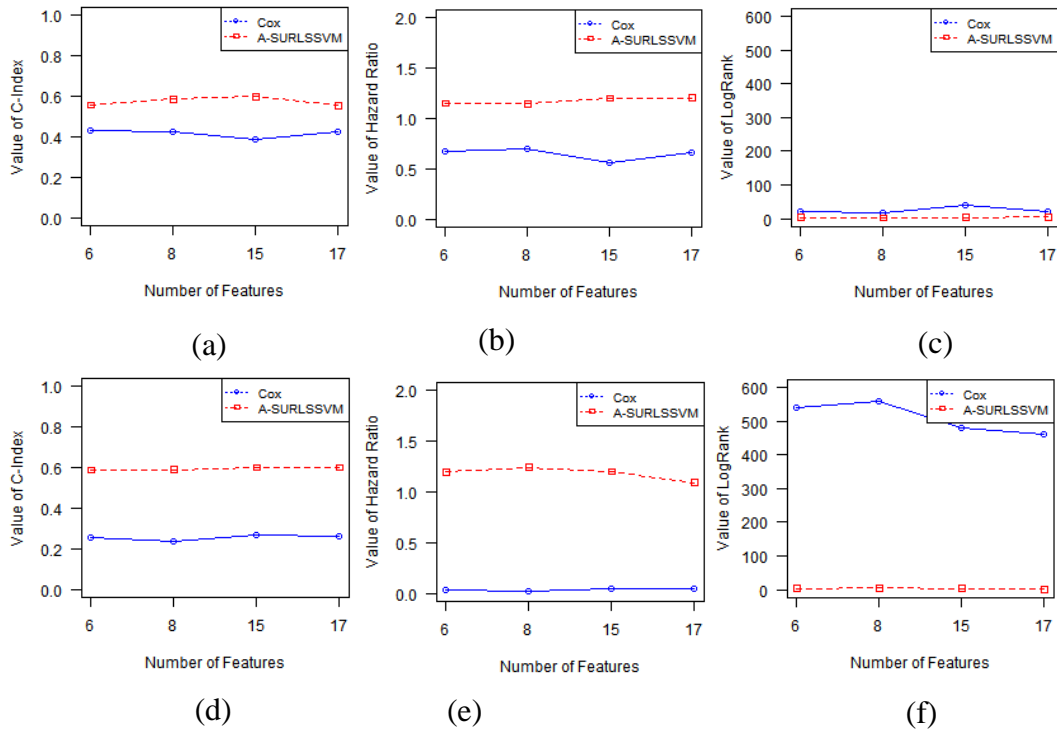
**Gambar 4.26** (a), (b), (c) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor 17 Jumlah Data 1000  
 (d), (e), (f) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor 17 Jumlah Data 1000

#### 4.3.6 Penerapan dengan Jumlah Variabel yang Berbeda pada Data dengan pola linier dan Nonlinier

Berikut adalah perbandingan performansi data dengan pola linier dan nonlinier antara metode A-SURLSSVM dan Cox PHM pada data dengan jumlah variabel 6,8,15 serta 17 dan jumlah data 1000 pada data dengan pola linier maupun nonlinier dengan persentase sensoring 50%.

Berdasarkan ketiga ukuran performansi yang digunakan dapat diketahui bahwa dengan menggunakan C-index dan Hazard Ratio, metode A-SURLSSVM lebih baik dari pada Cox PHM karena nilainya lebih tinggi baik pada data dengan pola linier maupun nonlinier. Dengan menggunakan A-SURLSSVM, pada data dengan pola linier, nilai C-index dan Hazard Ratio hampir sama dengan nonlinier. Namun, pada model Cox PHM nilai C-index dan Hazard Ratio selisihnya cukup besar antara data dengan pola linier dan nonlinier. C-index dan Hazard Ratio dari

Cox PHM pada data dengan pola nonlinier lebih rendah dari pada data dengan pola linier. Sehingga pada data dengan pola nonlinier performansi dari A-SURLSSVM jauh lebih bagus dibandingkan Cox PHM. Sebaliknya, jika dilihat dari Logrank, nilainya Logrank cox pada data dengan pola nonlinier jauh lebih tinggi dibandingkan A-SURLSSVM. Hasil yang sama juga didapatkan pada jumlah data yang lain (100 dan 5000) yang dapat dilihat pada lampiran.



**Gambar 4.27** (a), (b), (c) : Performansi pada Data dengan pola linier dengan Jumlah Variabel Prediktor Berbeda Jumlah Data 1000

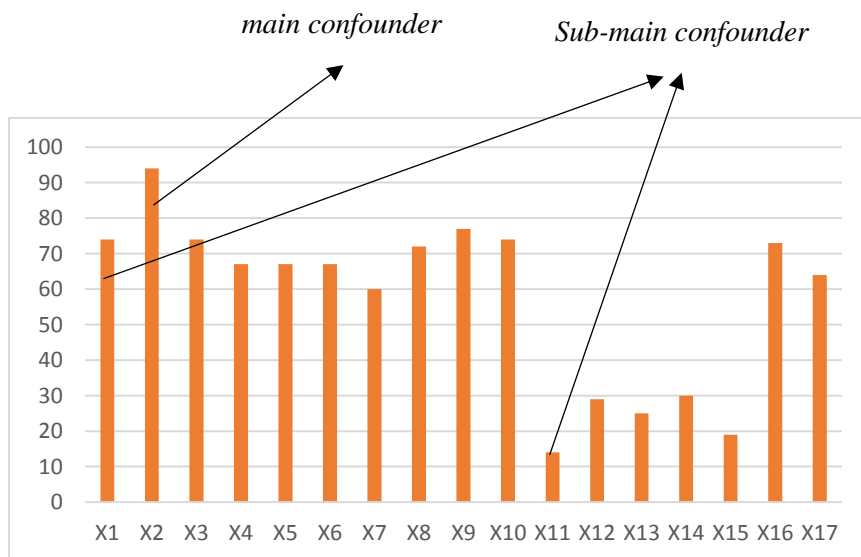
(d), (e), (f) : Performansi pada Data dengan pola nonlinier dengan Jumlah Variabel Prediktor Berbeda Jumlah Data 1000

Dari berbagai skenario yang telah dilakukan pada data simulasi dengan jumlah variabel, persentase sensing, dan jumlah data yang berbeda-beda, performansi dari A-SURLSSVM lebih baik dari Cox PHM berdasarkan C-index dan hazard ratio baik pada data dengan pola linier maupun nonlinier. Pada data dengan pola nonlinier selisih performansi C-index antara A-SURLSSVM dan Cox PHM lebih besar dibandingkan pada data dengan pola linier. Persentase sensing juga mempengaruhi performansi dari A-SURLSSVM. Semakin tinggi persentase sensing, maka performansi dari A-SURLSSVM semakin bagus. Sedangkan penambahan jumlah data dan variabel prediktor tidak terlalu berpengaruh terhadap performansi A-SURLSSVM. Sehingga metode A-SURLSSVM cocok digunakan

pada data dengan pola nonlinier dan data yang persentase sensoringnya besar. Selain itu, C-index merupakan ukuran performansi yang paling bagus dibandingkan hazard ratio dan log rank.

#### 4.3.7 Penerapan Metode A-SURLSSVM dengan *Features Selection* pada Data Simulasi

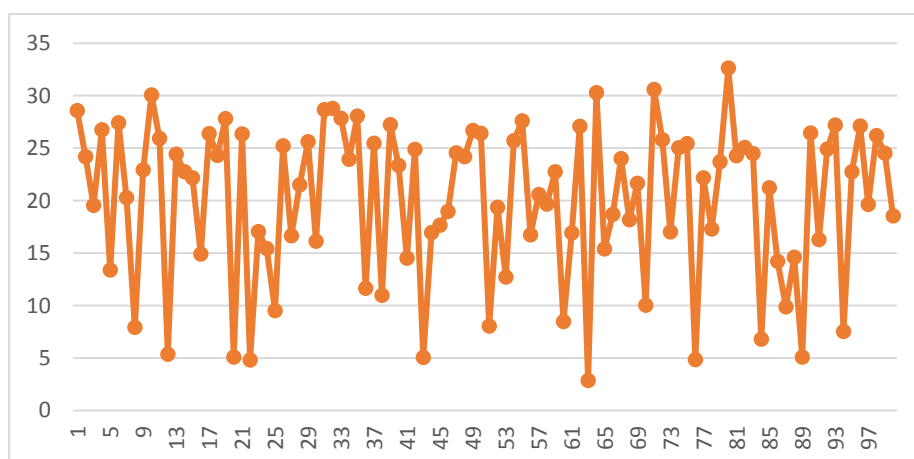
Penerapan *features Selection* dengan metode *Backward Elimination* dilakukan pada data simulasi dengan jumlah data sebanyak 1000 dan jumlah variabel 17 dimana kovariat pada  $X_{16}$  dan  $X_{17}$  disetting nol dan *survival time* dibangkitkan dalam bentuk nonlinier (terdapat interaksi) yaitu antara  $X_1$   $X_2$  dan  $X_2$   $X_{11}$ . Berikut adalah prosentase setiap variabel yang terpilih (tetap ada dalam model) dari 100 kali replikasi.



**Gambar 4.28** Prosentase Setiap Variabel yang Tetap Ada dalam Model dengan 100 kali Replikasi

Berdasarkan Gambar 4.28 dapat diketahui bahwa dari 100 kali replikasi, setiap variabel dari  $X_1$  hingga  $X_{17}$  pernah dihilangkan atau tetap berada di dalam model. Variabel yang paling besar prosentase tetap bertahan dalam model adalah  $X_2$  sebesar 94%. Sedangkan variabel yang paling sering dikeluarkan dalam model akibat *features selection* adalah  $X_{11}$  sebesar 14%. Sedangkan pada  $X_{16}$  dan  $X_{17}$  yang kovariatnya disetting nol, masih memiliki prosentase tetap dalam model yang cukup tinggi. Hal ini disebabkan beberapa hal, yang pertama *survival time* dibangkitkan

dari pola yang nonlinear dimana terdapat penambahan kovariat  $X_1 X_2$  dan  $X_2 X_{11}$ . Sedangkan model hanya menggunakan  $X_1$  hingga  $X_{17}$  (tanpa memasukkan kovariat yang nonlinear). Penyebab kedua adalah interaksi yang digunakan untuk membangkitkan *survival time* adalah  $X_1 X_2$  dan  $X_2 X_{11}$  dimana kedua interaksi menggunakan  $X_2$  sehingga  $X_2$  merupakan *main counfounder*. Hal inilah yang menyebabkan ketika  $X_{11}$  sebagai *sub-main confounder* dihilangkan dari model maka probabilitas  $X_2$  untuk dihilangkan dari model semakin kecil. Sama halnya dengan  $X_1$  sebagai *sub-main confounder*, prosentase variabel tersebut tetap bertahan dalam model akan lebih kecil dibandingkan  $X_2$  yang merupakan *main counfounder*.



**Gambar 4.29** Prosentase Kenaikan C-index setelah *Features Selection* dari 100 kali Replikasi

Berdasarkan grafik di atas, dapat diketahui prosentase kenaikan C-index sebelum hingga sesudah dilakukan *features selection* dari replikasi pertama hingga replikasi ke-100. Dengan menggunakan *features selection* dapat meningkatkan prosentase C-index hingga lebih dari 30%. Namun, juga terdapat beberapa replikasi yang prosentase kenaikan C-indexnya kurang dari 10%.

Dengan menggunakan *features selection*, dapat meningkatkan performansi dari A-SURLSSVM berdasarkan nilai C-index. Dari 100 replikasi yang telah dilakukan, didapatkan hasil sebagai berikut:



**Tabel 4.24** Nilai C-index dari 100 Replikasi

C-index	sebelum <i>feature selection</i>	setelah <i>feature selection</i>	Kenaikan C-index
Maksimum	57,2472	74,8612	23,6861
Minimum	48,9079	53,4480	1,56061
Rata-rata	52,6123	66,3762	13,7638

Berdasarkan Tabel 4.24 dapat diketahui bahwa dengan *features selection*, kenaikan C-index antara 1% hingga hampir 24%. Dari 100 replikasi, *features selection* dapat meningkatkan C-index dengan kenaikan rata-rata sekitar 13%. Sehingga nilai C-index yang awalnya sekitar 50an% meningkat hingga 70an%.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka didapatkan kesimpulan sebagai berikut:

1. Pada berbagai setting pada data simulasi dengan jumlah variabel, persentase sensoring, dan jumlah data yang berbeda-beda, performansi dari A-SURLSSVM lebih baik dari Cox PHM berdasarkan C-index dan hazard ratio baik pada data dengan pola linier maupun nonlinier. Pada data dengan pola nonlinier selisih performansi C-index antara A-SURLSSVM dan Cox PHM lebih besar dibandingkan pada data dengan pola linier. Persentase sensoring juga mempengaruhi performansi dari A-SURLSSVM. Semakin tinggi persentase sensoring, maka performansi dari A-SURLSSVM semakin bagus. Sedangkan penambahan jumlah data dan variabel prediktor tidak terlalu berpengaruh terhadap performansi A-SURLSSVM. Sehingga metode A-SURLSSVM cocok digunakan pada data dengan pola nonlinier dan data yang persentase sensoringnya besar. Selain itu, C-index merupakan ukuran performansi yang paling bagus dibandingkan hazard ratio dan log rank.
2. Hasil *features selection* pada data simulasi sebanyak 100 replikasi, secara umum dapat meningkatkan C-index secara signifikan. Selain itu, interaksi antar variabel menghasilkan variabel *main confounder* (yang paling besar probabilitasnya untuk tetap bertahan dalam model) dan *sub-main confounder* (variabel yang paling sering dikeluarkan dari model).
3. Pada data kanker serviks performansi A-SURLSSVM lebih baik dari pada Cox PHM berdasarkan C-index dan hazard ratio. Dengan *features selection*, dua dari dua belas variabel yang digunakan dikeluarkan dari model yaitu Komplikasi dan Riwayat KB sehingga dapat meningkatkan nilai C-index sebesar 1,4%.

## 5.2 Saran

Berdasarkan analisis dan pembahasan serta kesimpulan yang didapatkan, terdapat beberapa hal yang dapat menjadi rekomendasi baik untuk penelitian selanjutnya.

1. Dengan menggunakan data yang sama, dapat menggunakan metode lain seperti *survival* ANN untuk dibandingkan dengan A-SURLSSVM dan Cox PHM serta menggunakan C-index sebagai ukuran performansi karena C-index merupakan ukuran performansi yang paling cocok untuk data *survival*.
2. Pada penelitian ini *survival time* yang dibangkitkan dalam bentuk nonlinier (terdapat interaksi), namun tidak memasukkan data hasil interaksi ke dalam model sebagai variabel baru sehingga menghasilkan *features selection* yang kurang bagus. Untuk penelitian ke depannya, dapat memasukkan interaksi sebagai variabel baru ketika *survival time* yang dibangkitkan dalam bentuk nonlinier (terdapat interaksi).
3. Pada penelitian ini menggunakan metode *backward elimination* untuk *features selection* dan menghasilkan hasil yang kurang sesuai dengan teori. Sehingga untuk penelitian selanjutnya dapat menggunakan metode *features selection* lainnya.
4. Menggunakan data testing dan training serta cross validation sehingga menghasilkan akurasi yang lebih baik. Selain itu, dapat menggunakan metode optimasi parameter yang lain seperti PSO, GA dan metode-metode optimasi lainnya.
5. Pada penelitian selanjutnya memperhatikan kecukupan data pada masing-masing variabel kategori sehingga memenuhi syarat untuk dilakukan *crosstabulation*. Maka perlu dilakukan penggabungan data kategori pada variabel stadium dan pendidikan atau penambahan data untuk memenuhi kecukupan data.

## DAFTAR PUSTAKA

- Bakker, B., Heskes, T., Neijt, J., Kappen, B., (2004), "Improving Cox Survival Analysis with a Neural-Bayesian Approach", *Statistics in Medicine* 2004, Vol. 23, No. 19, 2989-3012, DOI:10.1002/sim.1904.
- Bender, Ralf., Augustin, Thomas., & Blettner, Maria, (2005), "Generating Survival Times to Simulate Cox Proportional Hazard Models", *Statistics in Medicine* 2005, 24:1713-1723.
- Bhavsar, Hetal, & A.Ganatra. (2012), "Variation of Support Vector Machine Classification Technique", *A Survey International Journal of Advanced Computer Research* (ISSN (print) : 2249-7277. ISSN (online) : 2277-7970), Vol. 2, No. 4, Issue 6 Desember 2012.
- Biganzoli, E., Boracehi, P., Marianti, L., Marubini, E., (1998), "Feedforward Neural Networks for the Analysis of Censored Survival Data : A Partial Logistic Regression Approach", *Statistics in Medicine*, Vol. 17, No. 10, 1169-1186.
- Boyd, S. & Vanderberghe, L., (2004), *Convex Optimization*, Cambridge University Press, Cambridgeshire, DOI:10.2277/0521833787.
- Chandrashekar, G. & Sahin, F. "A Survey on Feature Selection Methods", *Computers and Electrical Engineering*, Vol. 40, 16-28.
- Collett, D. (2003). *Modelling Survival Data in Medical Research* (2nd ed), Chapman and Hall, London.
- Cox, D. R., (1972), "Regression Models and Life-Tables (with Discussion)", *Journal of the Royal Statistical Society, Series B*, Vol. 34, No. 2, 187-220.
- Dalimartha, S., (2004), *Deteksi Dini Kanker dan Simplisia Anti Kanker*, Penebar Swadaya, Jakarta.
- Dizon, D. S., Krycman, M. L., & DiSivetro, P. A., (2011), "100 Tanya Jawab Mengenai Kanker Serviks" dalam *Analisis Survival pada Data Rekurensi dengan Pendekatan Andersen/Gill (Studi Kasus : Data Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya*, Wulansari, N. W., Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.

- Faraggi, D., & Simon, (1995), "A Neural Network Model for Survival Data", *Statistics in Medicine* 1995, Vol. 14, No. 1, 73-82.
- Gorunescu, F., (2011), *Intelligent Systems Reference Library*, (Gorunescu, Ed.).
- Gray, R. J., (1992), "Flexible Methods for Analyzing Survival Data Using Splines with Applications to Breast Cancer Prognosis", *Journal of the American Statistical Association* 87, 942-951.
- Haerdle, W. K., Prastyo, D. D., & Hafner, M., (2014), *Support Vector Machines with Evolutionary Model Selection for Default Prediction*, in J. Racine, L. Su, and A. Ullah (Eds.), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, New York, pp. 346-373.
- Harrell, F. Jr., Klee, K., Califf, R., Pryor, D., & Rosati, R., (1984), "Regression Modeling Strategies for Improved Prognostic Prediction", *Statistics in Medicine* 1984, Vol. 3, No. 2, 143-152, DOI:10.1002/sim.4780030207.
- Harrell, F. E., Lee, K. L., & Pollock, B. G., (1988), "Regression Models in Clinical Studies : Determining Relationships Between Predictors and Response", *Journal of the National Cancer Institute*, Vol. 80, No. 15, 1198-1202.
- Hastie, T., & Tibshirani, R., (1990), *Generalized Additive Models*, Chapman and Hall, New York.
- Hosmer, D. W., Lemeshow, S., & May, S., (2008), *Applied Survival Analysis : Regression Modelling of Time to Event Data (2nd ed)*, John Wiley and Sons. Inc, New York.
- Hsu, C., Chang, C., & Lin, C., (2004), *A Practical Guide to Support Vector Classification*, Information Engineering Taiwan University, Taipei.
- Hsu, C., Chang, C., & Lin, C., (2010), *A Practical Guide to Support Vector Classification*, Information Engineering Taiwan University, Taipei.
- Inayati, K. D., & Purnami, S. W., (2015), "Analisis Survival Nonparametrik Pada Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya Menggunakan Metode Kaplan Meier dan Uji Log Rank", *Jurnal Sains dan Seni ITS*, Institut Teknologi Sepuluh Nopember, Surabaya.
- Kementrian Kesehatan RI, (2015), *Panduan Pelayanan Klinik Kanker Serviks*, Komisi Penanggulangan Kanker Nasional.

- Kleinbaum, D. G., & Klein, M., (2012), *Survival Analysis: A Self Learning Text* (Third ed.), Springer, London.
- Kriner, Monika, (2007), *Survival Analysis with Multivariate Adaptive Regression Spline*, Disertation, Munchen University.
- Le, C. T., (1997), *Applied Survival Analysis*, John Wiley and Sons. Inc, New York.
- Lee, E. T., & Wang, J. W., (2003), *Statistical Methods for Survival Data Analysis* (3rd ed.), John Wiley and Sons Inc, New York.
- Li, H., & Luan, Y., (2003), “Kernel Cox Regression Models for Linking Gene Expression Profiles to Censored Survival Data”, *Pacific Symposium on Biocomputing* 2003, Vol. 8, 65-76, DOI:10.1142/979812776303\_0007.
- Li, H., & Luan, Y., (2005), “Boosting Proportional Hazards Models Using Smoothing Splines with Applications to High-Dimensional Microarray Data”, *Bioinformatics* 2005, Vol. 21, No. 10, 2403-2409, DOI : 10.1093/bioinformatics/bti324.
- Mahjub, H., Goli, S., Faradmal, J., & Soltanian, A. R., (2016), “Performance Evaluation of Support Vector Regression Models for Survival Analysis : A Simulation Study”, (IJACSA) *International Journal of Advanced Computer Science an Application*, Vol. 7, No 6.
- McCormick, C. C., & Giuntoli, R. L., (2011), “Panduan untuk Penderita Kanker Serviks”, dalam *Analisis Survival pada Data Rekurensi dengan Pendekatan Andersen/Gill (Studi Kasus : Data Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya*, Wulansari, N. W., Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.
- Melva, (2008), *Faktor-Faktor yang Mempengaruhi Kejadian Kanker Leher Rahim pada Penderita yang Datang Berobat di RSUP H. Adam Malik Medan Tahun 2008*, Tesis (Dipublikasikan), Universitas Sumatra Utara, Medan.
- Mercer, J., (1909), “Foundations of Positive and Negative Type, and Their connection with the Theory of Integral Equations”, *Philosophical Transactions of the Royal Society of London*, Vol. 25, 3-23.
- Murniati, & Lisuwarni, (2014, Februari-Juli), “Hubungan Pengetahuan dan Tingkat ekonomi dengan Perilaku Deteksi Dini Kanker Serviks Menggunakan Metode IVA”, *Jurnal Delima Harapan*, 2(1), 41-44.

- National Cancer Institute, (2014), *Understanding Prognosis Factors and Statistics About Survival*, diakses tanggal 18 Maret 2017 di <http://www.cancer.gov/cancertopics/diagnosisstaging/prognosis>
- Novitasari, D. A., (2014), *Analisis Survival pada Data Rekurensi dengan Menggunakan Counting Process Approach dan Model PWP-GT Study Kasus: Data Kanker Serviks, Tesis (Tidak Dipublikasikan)*, Institut Teknologi Sepuluh Nopember, Surabaya.
- Nugrahani, R. M., (2010), *Analisis Faktor-Faktor yang Mempengaruhi Hasil Pap Test Kanker Serviks dengan Menggunakan Metode Regresi Logistik Ordinal, Tugas Akhir*, Institut Teknologi Sepuluh Nopember, Surabaya.
- Nugroho, A. S., (2003), “Support Vector Machine Teori dan Aplikasinya dalam BioInformatika”, Ilmu Komputer.com.
- Ocviyanti, D., & Handoko, Y., (2013, Januari), “Peran Dokter Umum dalam Pencegahan Kanker Serviks di Indonesia”, *J Indon Med Assoc*, Vol. 63, No. 1, 1-3.
- Respita, R. D., (2017), *Perbandingan Model Regresi Spline dan Model Multivariate Adaptive Regression Splines untuk Analisis Survival pada Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya, Tesis (Tidak Dipublikasikan)*, Institut Teknologi Sepuluh Nopember, Surabaya.
- Robandi, I., & Prasetyo, G. R. A., (2008), *Peramalan Beban Jangka Pendek untuk Hari-Hari Libur dengan Metode Support Vector Machine, Tugas Akhir*, Institut Teknologi Sepuluh Nopember, Surabaya.
- Santosa, B., (2007), “Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis” dalam *Combine Sampling -Least Square Support Vector Machine untuk Klasifikasi Multi Class Imbalanced Data*, Khaulasari, H., Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.
- Schoenfeld, D., (1982, April), “Partial Residual for Proportional Hazard Regression Model”, *Biometrika*, Vol. 69, No. 1, 239-241.
- Scholkopf, B., & Smola, (2002), *Learning with Kernel : Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, Cambridge.

- Society, A. C., (2014), *Cervical Cancer Prevention and Early Detection*, American Cancer Society, Atlanta, Ga.
- Sullivan, F., (1988), "Nonparametric Estimation of Relative Risk Using Splines and Crossvalidation", *SIAM Journal on Scientific and Statistical Computing*, Vol. 9, 531-542. DOI:10.1137/0909035.
- Suykens, J. A. K., & Vandewalle, J., (1999a), "Training Multilayer Perceptron Classifiers Based on a Modified Support Vector Method". *IEEE Transactions on Neural Networks*, Vol. 10, 907-912.
- Suykens, J. A. K., & Vandewalle, J., (1999b), "Least Squares Support Vector Machine Classifiers". *Neural Processing Letters*, Vol. 9, 293-300.
- Suykens, J. A. K., & Vandewalle, J., (1999c), "Multiclass Least Squares Support Vector Machines". In Proc. Of the Int. Joint Conf. On Neural Networks (IJCNN'99), Washington, DC.
- Suykens, J. A. K., Van Gestel, T., De Brabanter, J., De Moor, B., & Vandewalle, J., (2002), *Least Squares Support Vector Machines*, World Scientific, Singapore.
- Trapsilasiwi, R. K., (2013), *Klasifikasi Multiclass untuk Imbalanced Data Menggunakan SMOTE Least Square Support Vector Machine*, Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.
- Van Belle, V. Pleckmans, K., Suykens, J. A., & Van Huffel, S., (2007), "Support Vector Machines for Survival Analysis", *Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007)*, Plymouth (UK), 1-8.
- Van Belle, V. Pleckmans, K., Suykens, J. A., & Van Huffel, S., (2008), "Survival SVM : a Practical Scalable Algorithm", *Proceedings of the 16th European Symposium on Artificial Neural Networks (ESANN2008)*, Bruges (Belgium), 89-94.
- Van Belle, V. Pleckmans, K., Suykens, J. A., & Van Huffel, S., (2010), "Additive Survival Least Square Support Vector Machines", *Statistics in Medicine*, Vol. 29, No. 2, 296-308.



- Van Belle, V. Pleckmans, K., Suykens, J. A., & Van Huffel, S., (2011), “Support Vector Methods for Survival Analysis : A Comparison Between Ranking and Regression Approaches”, *Artificial Intelligence in Medicine*, 53, 107-118.
- Vapnik, V., (1995), “The Nature of Statistical Learning Theory”, Springer-Verlag, New York.
- WHO, I. A., (2005), *Cervix Cancer Screening* (Vol. 10), IARCH Press, Lyon.
- Wijayanti, R., (2014), *Perbandingan Analisis Regresi Cox dan Analisis Survival Bayesian pada Penderita Kanker serviks di RSUD Dr. Soetomo Surabaya*, Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.
- Wulansari, N. W., (2015), *Analisis Survival pada Data Rekurensi dengan Pendekatan Andersen/Gill (Studi Kasus : Data Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya*, Tesis (Tidak Dipublikasikan), Institut Teknologi Sepuluh Nopember, Surabaya.

**Lampiran 1. Surat Keterangan Kelayakan Etik**

F.LITB.003



**KOMITE ETIK PENELITIAN KESEHATAN  
RSUD Dr. SOETOMO SURABAYA**

**KETERANGAN KELAIKAN ETIK  
("ETHICAL CLEARANCE")**

376 / Panke.KKE / V / 2017

**KOMITE ETIK RSUD Dr. SOETOMO SURABAYA TELAH MEMPELAJARI SECARA SEKSAMA RANCANGAN PENELITIAN YANG DIUSULKAN, MAKA DENGAN INI MENYATAKAN BAHWA PENELITIAN DENGAN JUDUL :**

**" Addictive Survival Least Square SVM untuk Analisis Data Survival  
(Studi Simulasi dan Studi Kasus pada Data Pasien Kanker Serviks di RSUD Dr. Soetomo Surabaya) "**

**PENELITI UTAMA : Chusnul Khotimah**

**PENELITI LAIN : 1. Dr. Santi Wulan Purnami, S.Si., M.Si  
2. Dr. rer. Pol. Dedy Dwi Prasetyo, S.Si., M.Si**

**UNIT / LEMBAGA / TEMPAT PENELITIAN: RSUD Dr. Soetomo Surabaya**

**DINYATAKAN LAIK ETIK**

**SURABAYA, 30 MAY 2017**



**(Dr. Elizeus Hanindito, dr., Sp.An, KIC,KAP)  
NIP. 19511007 197903 1 002**

**Lampiran 2.** Syntax Membangkitkan Data Simulasi untuk  $p=6$  pada Data dengan Pola Linier

```
library(multinomRob)
#banyak data
n=100
#n=1000
#n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rmultinom(n, size=1,prob = c(0.5,0.2,0.2,0.1))
x3=t(x3)
x3dummy=x3[,-1]
x31=x3dummy[,1]
x32=x3dummy[,2]
x33=x3dummy[,3]
x4=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x4=t(x4)
x4dummy=x4[,-1]
x41=x4dummy[,1]
x42=x4dummy[,2]
x5=rnorm(n,40,2)
x6=rnorm(n,20,3)
x=cbind(x1,x2,x31,x32,x33,x41,x42,x5,x6)
#b adalah besar parameter beta
b=(c(0.01,-0.15 ,-0.07,0.035,-0.8,0.0015, -0.075,-0.075, -0.02))
fix 1b=(c(0.01,-0.15 ,-0.07,0.035,-0.8,0.0015, -0.075,-0.075, -0.02))
xb=x%*%b
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,0.5)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6)
library(data.table)
write.table(data,file="D:/data1.csv",sep=",")
write.table(data,file="D:/data1.txt")
require(vcd)
require(MASS)
data=read.table("D:/data1.txt")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
```

### Lampiran 3. Syntax Membangkitkan Data Simulasi untuk $p=8$ pada Data dengan Pola Linier

```
library(multinomRob)
#banyak data
n=100
#n=1000
#n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rmultinom(n, size=1,prob = c(0.5,0.2,0.2,0.1))
x3=t(x3)
x3dummy=x3[,-1]
x31=x3dummy[,1]
x32=x3dummy[,2]
x33=x3dummy[,3]
x4=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x4=t(x4)
x4dummy=x4[,-1]
x41=x4dummy[,1]
x42=x4dummy[,2]
x5=rnorm(n,40,2)
x6=rnorm(n,20,3)
x7=rbinom(n,1,0.5)
x8=rnorm(n,0,1)
x=cbind(x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
#b adalah besar parameter beta
b=(c(0.01,-0.15,-0.07,0.035,-0.08,0.0015,-0.075,-0.085,-0.02,0,0))
xb=x%*%b
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,0.5)
t=ceiling(t)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
library(data.table)
write.table(data,file="D:/data2.csv",sep=",")
write.table(data,file="D:/data2.txt")
require(vcd)
require(MASS)
data=read.table("D:/data100.txt")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
```

**Lampiran 4.** Syntax Membangkitkan Data Simulasi untuk  $p=15$  pada Data dengan Pola Linier

```
library(multinomRob)
#banyak data
#n=100
#n=1000
n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rbinom(n,1,0.5)
x4=rbinom(n,1,0.6)
x5=rbinom(n,1,0.7)
x6=rmultinom(n, size=1,prob = c(0.5,0.3,0.1,0.1))
x6=t(x6)
x6dummy=x6[,-1]
x61=x6dummy[,1]
x62=x6dummy[,2]
x63=x6dummy[,3]
x7=rmultinom(n, size =1, prob = c(0.05,0.4,0.5,0.05))
x7=t(x7)
x7dummy=x7[,-1]
x71=x7dummy[,1]
x72=x7dummy[,2]
x73=x7dummy[,3]
x8=rmultinom(n, size =1, prob = c(0.3,0.5,0.2))
x8=t(x8)
x8dummy=x8[,-1]
x81=x8dummy[,1]
x82=x8dummy[,2]
x9=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x9=t(x9)
x9dummy=x9[,-1]
x91=x9dummy[,1]
x92=x9dummy[,2]
x10=rmultinom(n, size =1, prob = c(0.05,0.3,0.6,0.05))
x10=t(x10)
x10dummy=x10[,-1]
x101=x10dummy[,1]
x102=x10dummy[,2]
x103=x10dummy[,3]
x11=rnorm(n,40,3)
x12=rnorm(n,40,2)
x13=rnorm(n,30,2)
x14=rnorm(n,20,2)
x15=rnorm(n,20,3)
x=cbind(x1,x2,x3,x4,x5,x61,x62,x63,x71,x72,x73,x81,x82,x91,x92,
x101,x102,x103,x11,x12,x13,x14,x15)
```

```

#b adalah besar parameter beta
b=(c(0.01,-0.15 , -0.12,0.012,0.008,-0.07,0.003,-0.08,0.015, -0.075,-0.07,0.035,-
0.06,0.025, -0.065,0.01,-0.06,0.03,-0.04, -0.015,-0.015,-0.01,-0.035))
xb=x%*%b
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,0.5)
t=ceiling(t)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
library(data.table)
write.table(data,file="D:/data3.csv",sep=",")
write.table(data,file="D:/data3.txt")
require(vcd)
require(MASS)
data=read.table("D:/data3.txt")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
statdes

```

**Lampiran 5.** Syntax Membangkitkan Data Simulasi untuk  $p=17$  pada Data dengan Pola Linier

```
library(multinomRob)
#banyak data
#n=100
#n=1000
n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rbinom(n,1,0.5)
x4=rbinom(n,1,0.6)
x5=rbinom(n,1,0.7)
x6=rmultinom(n, size=1,prob = c(0.5,0.3,0.1,0.1))
x6=t(x6)
x6dummy=x6[,-1]
x61=x6dummy[,1]
x62=x6dummy[,2]
x63=x6dummy[,3]
x7=rmultinom(n, size =1, prob = c(0.05,0.4,0.5,0.05))
x7=t(x7)
x7dummy=x7[,-1]
x71=x7dummy[,1]
x72=x7dummy[,2]
x73=x7dummy[,3]
x8=rmultinom(n, size =1, prob = c(0.3,0.5,0.2))
x8=t(x8)
x8dummy=x8[,-1]
x81=x8dummy[,1]
x82=x8dummy[,2]
x9=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x9=t(x9)
x9dummy=x9[,-1]
x91=x9dummy[,1]
x92=x9dummy[,2]
x10=rmultinom(n, size =1, prob = c(0.05,0.3,0.6,0.05))
x10=t(x10)
x10dummy=x10[,-1]
x101=x10dummy[,1]
x102=x10dummy[,2]
x103=x10dummy[,3]
x11=rnorm(n,40,3)
x12=rnorm(n,40,2)
x13=rnorm(n,30,2)
x14=rnorm(n,20,2)
x15=rnorm(n,20,3)
x16=rbinom(n,1,0.5)
x17=rnorm(n,0,1)
```

```

x=cbind(x1,x2,x3,x4,x5,x61,x62,x63,x71,x72,x73,x81,x82,x91,x92,
x101,x102,x103,x11,x12,x13,x14,x15,x16,x17)
#b adalah besar parameter beta
b=(c(0.01,-0.15 , -0.12,0.012,0.008,-0.07,0.003,-0.08,0.015, -0.075,-0.07,0.035,-
0.06,0.025, -0.065,0.01,-0.06,0.03,-0.04, -0.015,-0.015,-0.01,-0.035))
xb=x%*%b
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,0.5)
t=ceiling(t)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
library(data.table)
write.table(data,file="D:/data4.csv",sep=",")
write.table(data,file="D:/data4.txt")
require(vcd)
require(MASS)
data=read.table("D:/data100.txt")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
statdes

```



**Lampiran 6.** Syntax Membangkitkan Data Simulasi untuk  $p=6$  pada Data dengan Pola Nonlinier

```
library(multinomRob)
#banyak data
n=100
#n=1000
#n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rmultinom(n, size=1,prob = c(0.5,0.2,0.2,0.1))
x3=t(x3)
x3dummy=x3[,-1]
x31=x3dummy[,1]
x32=x3dummy[,2]
x33=x3dummy[,3]
x4=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x4=t(x4)
x4dummy=x4[,-1]
x41=x4dummy[,1]
x42=x4dummy[,2]
x5=rnorm(n,40,2)
x6=rnorm(n,20,3)
x=cbind(x1,x2,x31,x32,x33,x41,x42,x5,x6)
#b adalah besar parameter beta
b=(c(0.01,-0.15 , -0.07,0.035,-0.08,0.0015, -0.075,-0.085, -0.02))
xb=x%*%b
xb=xb+0.03*x1*x2+0.09*x2*x5
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,1)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6)
library(data.table)
write.table(data,file="D:/bismillah tesis/data simulasi/non/1 x=6 n=5000
non.csv",sep=",")
require(vcd)
require(MASS)
data=read.csv("D:/bismillah tesis/data simulasi/non/1 x=6 n=5000
non.csv",sep=",")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
statdes
```

### Lampiran 7. Syntax Membangkitkan Data Simulasi untuk $p=8$ pada Data dengan Pola Nonlinier

```
library(multinomRob)
#banyak data
n=100
#n=1000
#n=5000
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rmultinom(n, size=1,prob = c(0.5,0.2,0.2,0.1))
x3=t(x3)
x3dummy=x3[,-1]
x31=x3dummy[,1]
x32=x3dummy[,2]
x33=x3dummy[,3]
x4=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x4=t(x4)
x4dummy=x4[,-1]
x41=x4dummy[,1]
x42=x4dummy[,2]
x5=rnorm(n,40,2)
x6=rnorm(n,20,3)
x7=rbinom(n,1,0.5)
x8=rnorm(n,0,1)
x=cbind(x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
#b adalah besar parameter beta
b=(c(0.01,-0.15,-0.07,0.035,-0.08,0.0015,-0.075,-0.085,-0.02,0,0))
xb=x%*%b
xb=xb+0.03*x1*x2+0.09*x2*x5
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xb)))
#delta (notasi status tersensor)
delta=rbinom(n,1,1)
data=data.frame(t,delta,x1,x2,x31,x32,x33,x41,x42,x5,x6,x7,x8)
library(data.table)
write.table(data,file="D:/bismillah tesis/data simulasi/non/1 x=8 n=5000
non.csv",sep=",")
require(vcd)
require(MASS)
data=read.csv("D:/bismillah tesis/data simulasi/non/1 x=8 n=5000
non.csv",sep=",")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
```

**Lampiran 8.** Syntax Membangkitkan Data Simulasi untuk  $p=15$  pada Data dengan Pola Nonlinier

```
library(multinomRob)
#banyak data
n=100
#n=1000
#n=5000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rbinom(n,1,0.5)
x4=rbinom(n,1,0.6)
x5=rbinom(n,1,0.7)
x6=rmultinom(n, size=1,prob = c(0.5,0.3,0.1,0.1))
x6=t(x6)
x6dummy=x6[,-1]
x61=x6dummy[,1]
x62=x6dummy[,2]
x63=x6dummy[,3]
x7=rmultinom(n, size =1, prob = c(0.05,0.4,0.5,0.05))
x7=t(x7)
x7dummy=x7[,-1]
x71=x7dummy[,1]
x72=x7dummy[,2]
x73=x7dummy[,3]
x8=rmultinom(n, size =1, prob = c(0.3,0.5,0.2))
x8=t(x8)
x8dummy=x8[,-1]
x81=x8dummy[,1]
x82=x8dummy[,2]
x9=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x9=t(x9)
x9dummy=x9[,-1]
x91=x9dummy[,1]
x92=x9dummy[,2]
x10=rmultinom(n, size =1, prob = c(0.05,0.3,0.6,0.05))
x10=t(x10)
x10dummy=x10[,-1]
x101=x10dummy[,1]
x102=x10dummy[,2]
x103=x10dummy[,3]
x11=rnorm(n,40,3)
x12=rnorm(n,40,2)
x13=rnorm(n,30,2)
x14=rnorm(n,20,2)
x15=rnorm(n,20,3)
x=cbind(x1,x2,x3,x4,x5,x61,x62,x63,x71,x72,x73,x81,x82,x91,x92,
```

```

#b adalah besar parameter beta
b=(c(0.01,-0.15,-0.12,0.012,0.008,-0.07,0.003,-0.08,0.015,-0.075,-0.07,0.035,-
0.06,0.025,-0.065,0.01,-0.06,0.03,-0.04,-0.015,-0.015,-0.01,-0.035))
xb=x%*%b
xbt=xb+0.03*x1*x2+0.09*x2*x11
# tersimpan fix 1 lambda=1
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xbt)))
ceiling(t)

#delta (notasi status tersensor)
delta=rbinom(n,1,1)
data=data.frame(t,delta,x)
library(data.table)
write.table(data,file="D:/bismillah tesis/data simulasi/non/1 x=17 n=1000
non.csv",sep=",")

require(vcd)
require(MASS)
data=read.csv("D:/bismillah tesis/data simulasi/non/1 x=17 n=1000
non.csv",sep=",")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
statdes

```

**Lampiran 9.** Syntax Membangkitkan Data Simulasi untuk p=17 pada Data dengan Pola Nonlinier

```
library(multinomRob)
#banyak data
n=1000
#membangkitkan x
x1=rbinom(n,1,0.4)
x2=rbinom(n,1,0.3)
x3=rbinom(n,1,0.5)
x4=rbinom(n,1,0.6)
x5=rbinom(n,1,0.7)
x6=rmultinom(n, size=1,prob = c(0.5,0.3,0.1,0.1))
x6=t(x6)
x6dummy=x6[,-1]
x61=x6dummy[,1]
x62=x6dummy[,2]
x63=x6dummy[,3]
x7=rmultinom(n, size =1, prob = c(0.05,0.4,0.5,0.05))
x7=t(x7)
x7dummy=x7[,-1]
x71=x7dummy[,1]
x72=x7dummy[,2]
x73=x7dummy[,3]
x8=rmultinom(n, size =1, prob = c(0.3,0.5,0.2))
x8=t(x8)
x8dummy=x8[,-1]
x81=x8dummy[,1]
x82=x8dummy[,2]
x9=rmultinom(n, size =1, prob = c(0.4,0.4,0.2))
x9=t(x9)
x9dummy=x9[,-1]
x91=x9dummy[,1]
x92=x9dummy[,2]
x10=rmultinom(n, size =1, prob = c(0.05,0.3,0.6,0.05))
x10=t(x10)
x10dummy=x10[,-1]
x101=x10dummy[,1]
x102=x10dummy[,2]
x103=x10dummy[,3]
x11=rnorm(n,40,3)
x12=rnorm(n,40,2)
x13=rnorm(n,30,2)
x14=rnorm(n,20,2)
x15=rnorm(n,20,3)
x16=rbinom(n,1,0.5)
x17=rnorm(n,0,1)
```

```

x=cbind(x1,x2,x3,x4,x5,x61,x62,x63,x71,x72,x73,x81,x82,x91,x92,
x101,x102,x103,x11,x12,x13,x14,x15,x16,x17)
#b adalah besar parameter beta
b=(c(0.01,-0.15,-0.12,0.012,0.008,-0.07,0.003,-0.08,0.015,-0.075,-0.07,0.035,-
0.06,0.025,-0.065,0.01,-0.06,0.03,-0.04,-0.015,-0.015,-0.01,-0.035,0,0))
xb=x%*%b
xbt=xb+0.03*x1*x2+0.09*x2*x11
# tersimpan fix 1 lambda=1
lambda=1
u=runif(n, min=0, max=1)
#membangkitkan t (survival time)
t=-log(u)/(lambda*(exp(xbt)))
ceiling(t)

#delta (notasi status tersensor)
delta=rbinom(n,1,1)
data=data.frame(t,delta,x)
library(data.table)
write.table(data,file="D:/bismillah tesis/data simulasi/non/1 x=17 n=1000
non.csv",sep=",")

require(vcd)
require(MASS)
data=read.csv("D:/bismillah tesis/data simulasi/non/1 x=17 n=1000
non.csv",sep=",")
t=data$t
fit1 <- fitdistr(t, "exponential")
ks=ks.test(t, "pexp", fit1$estimate)
ks$p.value
statdes=cbind(min=min(t),max=max(t), mean=mean(t),median=median(t))
statdes

```

**Lampiran 10.** Data Pasien Kanker Serviks Rumah Sakit dr. Soetomo 2013-2016

No	T	Status	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>41</sub>	X <sub>42</sub>	X <sub>43</sub>	X <sub>51</sub>	X <sub>52</sub>	X <sub>53</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11(1)</sub>	X <sub>11(2)</sub>	X <sub>11(3)</sub>	X <sub>12(1)</sub>	X <sub>12(2)</sub>	X <sub>12(3)</sub>
1	712	0	43	0	1	0	0	0	1	0	0	20	14	28	7	3	1	0	0	0	0	1
2	625	0	49	1	1	0	0	1	0	1	0	18	13	30	7	2	0	1	0	0	0	1
3	505	0	32	1	1	0	0	1	0	1	0	23	12	28	7	3	1	0	0	1	0	0
4	361	0	48	1	1	0	0	1	0	0	1	18	14	28	5	2	1	0	0	1	0	0
5	254	1	39	1	1	0	0	0	0	1	0	18	12	30	7	4	1	0	0	0	0	1
6	277	0	37	1	1	0	0	1	0	1	0	20	12	28	4	2	0	1	0	0	1	0
7	66	0	58	1	0	1	0	0	0	0	1	16	15	28	7	2	0	1	0	1	0	0
8	170	0	53	1	1	0	0	1	0	0	1	21	14	28	7	3	1	0	0	1	0	0
9	105	0	78	0	0	1	0	0	0	1	0	18	12	28	7	5	0	0	1	1	0	0
10	50	0	52	1	0	1	0	0	0	1	0	23	16	30	5	2	0	0	1	1	0	0
11	64	0	50	1	1	1	0	0	0	0	1	21	13	28	7	4	0	0	1	0	0	1
12	325	0	48	1	1	0	0	1	0	0	1	17	15	28	5	2	0	0	0	1	0	0
13	116	0	39	0	1	0	0	1	0	1	0	18	12	28	7	4	0	1	0	0	0	1
14	74	0	49	1	0	1	0	0	0	0	1	17	13	30	10	3	0	0	1	0	1	0
15	131	0	54	1	0	1	0	0	0	1	0	20	12	28	7	3	0	0	0	1	0	0
16	155	0	54	1	1	0	0	1	0	1	0	19	13	28	7	2	0	1	0	0	0	1
17	80	0	51	1	1	0	0	1	0	0	1	13	12	28	7	2	1	0	0	0	1	0
18	287	0	46	0	1	0	0	1	0	0	1	23	14	28	7	2	0	0	0	0	0	1
19	107	0	37	1	0	1	0	0	0	1	0	25	12	28	7	2	1	0	0	0	0	1
20	97	0	49	1	1	0	0	1	0	1	0	18	15	28	7	2	0	0	0	0	0	1
21	70	0	46	1	1	1	0	0	0	0	1	22	13	28	6	2	1	0	0	0	0	1
22	198	0	47	1	1	0	0	1	0	1	0	19	15	30	7	2	0	0	0	1	0	0
23	57	0	57	1	1	0	0	1	0	0	1	19	17	28	5	1	0	0	0	1	0	0

**Lampiran 10.** Data Pasien Kanker Serviks Rumah Sakit dr. Soetomo 2013-2016 (Lanjutan)

No	T	Status	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>41</sub>	X <sub>42</sub>	X <sub>43</sub>	X <sub>51</sub>	X <sub>52</sub>	X <sub>53</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11(1)</sub>	X <sub>11(2)</sub>	X <sub>11(3)</sub>	X <sub>12(1)</sub>	X <sub>12(2)</sub>	X <sub>12(3)</sub>
24	205	0	61	1	0	1	0	0	0	0	1	16	14	30	4	5	1	0	0	0	0	1
25	87	0	51	0	0	1	0	0	0	0	1	13	12	28	7	3	0	1	0	0	0	1
26	79	0	32	1	1	0	0	1	0	1	0	19	14	28	7	2	0	1	0	1	0	0
27	65	0	65	1	0	1	0	0	0	0	1	15	12	30	7	4	1	0	0	1	0	0
28	27	0	51	0	0	1	0	0	0	1	0	24	13	28	10	3	0	1	0	0	0	1
29	59	0	34	1	0	1	0	0	0	1	0	24	14	28	7	1	0	0	1	0	0	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
406	154	0	40	0	1	0	0	1	0	1	0	22	12	28	7	0	1	0	0	0	0	1
407	67	0	45	0	0	1	0	0	0	0	1	20	14	28	7	2	0	1	0	0	0	0
408	62	0	34	0	1	0	0	1	1	0	0	19	11	28	7	3	0	1	0	0	0	1
409	46	0	47	0	0	1	0	0	0	1	0	19	12	28	7	5	1	0	0	0	1	0
410	68	0	55	0	0	1	0	0	0	0	1	20	14	28	7	3	1	0	0	0	1	0
411	114	0	29	1	1	0	1	0	0	1	0	23	14	28	7	3	1	0	0	0	1	0
412	72	0	48	0	0	1	0	0	0	1	0	17	13	28	7	3	0	1	0	1	0	0

**Keterangan :**

$t$	<i>Survival time</i>	$x_6$	Usia Saat Menikah
Status	0 = tidak terjadi <i>event</i>	$x_7$	Usia pertama kali menstruasi
	1= terjadi <i>event</i> (meninggal)	$x_8$	Siklus menstruasi
		$x_9$	Lama menstruasi



$x_1$	Usia Pasien	$x_{10}$	Frekuensi Melahirkan (Paritas)
	Komplikasi		Riwayat KB
	0 = Tidak Ada		0 = Tidak
$x_2$	1 = Ada	$x_{11}$	1 = Suntik
	Status Anemia		2 = Pil
	0 = Tidak		3 = IUD
$x_3$	1 = Ya	$x_{12}$	Pendidikan
	Jenis Pengobatan		0 = SD
	0 = Kemoterapi		1 = SMP
$x_4$	1 = Transfusi PRC		2 = SMA
	2 = Keduanya		3 = Sarjana
	3 = Lainnya		
$x_5$	Stadium		
	1 = Stadium I		
	2 = Stadium II		
	3 = Stadium III		
	4 = Stadium IV		

### Lampiran 11. Syntax untuk Menghitung Matriks D

```
# Y dan delta
makeDiff = function(Y,delta) {
  nb  = 1
  n   = length(Y)
  si  = order(Y)
  Ybaru = Y[si]
  delta = delta[si]
  e1   = min(which(delta==1))
  ne   = n - e1
  D    = matrix(0,ne,n)
  diffx = rep(0,)
  t    = 1
  eve  = e1
  for (i in 1:ne) {
    ada = e1 + i
    b2  = 0
    eve = max(which(delta[1:ada-1]==1))
    # print(eve)
    D[t,eve] = -1
    D[t,ada] = 1;
    diffx[t] = Ybaru[ada] - Ybaru[eve]
    # print(diffx)
    b2 = b2 + 1
    t  = t + 1
    while (b2<nb & (eve-1)>0) {
      eve = max(which(delta[1:eve-1]==1))
      if (length(eve)==0) {
        break
      }
    }

    # print(eve)

    D[t,eve] = -1
    D[t,ada] = 1
    diffx[t] = Ybaru[ada] - Ybaru[eve]

    # print(diffx)

    b2    = b2 + 1
    t    = t + 1
  }
}
Dx = D[,si]
return(Dx)
}
```

## Lampiran 12. Syntax untuk Menghitung Nilai Fungsi Kernel

```
kernel_rbf = function(X,kern_par) {  
  X = X  
  X = as.matrix(X)  
  kpar = kern_par  
  
  nr = dim(X)[1]  
  nc = dim(X)[2]  
  
  XXh = matrix(rep(apply(X*X,1,sum),nr),ncol = nr)  
  omega = XXh+t(XXh) - 2*(X%*%t(X))  
  omega = exp(-omega/(2*kpar))  
  
  return(omega)  
}
```

### Lampiran 13. Syntax untuk Cox PHM dan Ukuran Performansinya

```
#memanggil data
data=read.csv("D:/bismillah    tesis/data    simulasi/1    x=17    n=1000
linier.csv",sep=",")
library(survival)
data=data.frame(data)
attach (data)
local ({pkg=select.list(sort(.packages(all.available=TRUE)))
      if(nchar(pkg))library(pkg,character.only=TRUE)})
#estimasi parameter Cox PHM
library (splines)
x=cbind(x1,x2,x3,x4,x5,x61,x62,x63,x71,x72,x73,x81,x82,x91,x92,
x101,x102,x103,x11,x12,x13,x14,x15,x16,x17)
coxph=coxph(Surv(ceiling(t),delta)~x,method="efron")
coxph
koefisien=coxph$coef
koefisien
prognostik_indeks=x%*%koefisien

#hazardratio (hr)
prog=prognostik_indeks
library(foreign)
medprog=mean(prognostik_indeks)
n=1000
for (i in 1:n){
if(prog[i] > medprog)
prog[i]=1 else prog[i]=2}
coxph_hr=coxph(Surv(ceiling(t),delta)~prog,method="efron")
coxph_hr
coef_hr=coxph_hr$coef
hr=exp(coef_hr)

#logrank
prog=prognostik_indeks
library(foreign)
medprog=mean(prognostik_indeks)
for (i in 1:n){
if(prog[i] > medprog)
prog[i]=1 else prog[i]=2}
prog
logrank=survdiff(Surv(ceiling(t),delta) ~prog)
```

### Lampiran 13. Syntax untuk Cox PHM dan Ukuran Performansinya (Lanjutan)

```
#concordanceindeks
memory.limit(size=10000)
library (combinat)
require(combinat)
comprog=combn(prognostik_indeks, 2)
comprog=t(comprog)
prog_i=cbind(comprog[,1])
prog_j=cbind(comprog[,2])
selisih_prog=prog_j-prog_i
comtime=combn(t, 2)
comtime=t(comtime)
ti=cbind(comtime[,1])
tj=cbind(comtime[,2])
selisih_t=tj-ti
indikator=selisih_prog*selisih_t
for (i in 1:length(indikator)){
  if(indikator[i] < 0)
  indikator[i]=0 else indikator[i]=1 }
comstatus=combn(delta, 2)
comstatus=t(comstatus)
status_i=cbind(comstatus[,1])
v=matrix(0,nrow=length(selisih_t),ncol=1)
for (i in 1:length(selisih_t)){
  if((selisih_t[i]>0) & status_i[i]==1)
  v[i]=1 else v[i]=0}
cindeks=sum(indikator*v)/sum(v)
hr
logrank
cindeks
```

#### Lampiran 14. Syntax untuk A-SURLSSVM dan Ukuran Performansinya

```
Perf.surls = data.frame(mu = c(), C = c(), kern_par = c(),
                        HR = c(), Logrank = c(),
                        C.index=c())
Perbandingan = data.frame(HR.SURLS = c(),
                          Logrank.SURLS = c(),
                          C.index.SURLS = c(),
                          mu = c(), C = c(), kern_par = c())
n = 412

data=read.csv("D://bismillah tesis/data serviks.csv",sep=";")
data
head(data)
x1=data$x1
x2=data$x2
x3=data$x3
x41=data$x41
x42=data$x42
x43=data$x43
x51=data$x51
x52=data$x52
x53=data$x53
x6=data$x6
x7=data$x7
x8=data$x8
x9=data$x9
x10=data$x10
x111=data$x111
x112=data$x112
x113=data$x113
x121=data$x121
x122=data$x122
x123=data$x123
X =
data.frame(x1,x2,x3,x41,x42,x43,x51,x52,x53,x6,x7,x8,x9,x10,x111,x112,x113
,x121,x122,x123)
Y = data$T
delta = data$Status
Time=data$T
library(survival)
library(splines)

mu = c(10^(-3),0.01,0.1,1,10)
C = c(0.01,0.1,1,10,100)
kern_par = c(0.5,0.25,1,2,4)
```

**Lampiran 14.** Syntax untuk A-SURLSSVM dan Ukuran Performansinya  
(Lanjutan)

```
for (ci in 1:length(C)) {
  for (ki in 1:length(kern_par)) {
    for (mui in 1:length(mu)) {

      print(paste("Estimasi A-SURLS SVM dengan mu =", mu[mui],
                  "C =", C[ci], "dan Kernel Parameter =", kern_par[ki]))

      source("./asurlssvm.R")
      surls = asurlssvm(Y,X,delta,C[ci],kern_par[ki],mu[mui])
      p.surls = surls$prognostik_indeks
      p.surls = as.matrix(p.surls)

      p.surls.med = mean(p.surls)
      p.surls.hasil = rep(NA,n)

      #Determined group by A-SURLS-SVM
      for (m in 1:n)
      {
        if (p.surls[m]>p.surls.med) # 1:high risk
          {p.surls.hasil[m] = 1}
        else
          {p.surls.hasil[m] = 0} # 0 : low risk
      }

      g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic A-SURLS-SVM

      # HAZARD RATIO
      reg.HR = coxph(Surv(Time,delta)~g.surls,method="efron")
      coeff.HR = reg.HR$coef
      HR = exp(coeff.HR)

      # LOG-RANK
      Logrank = survdiff(Surv(Time,delta)~g.surls)

      # Concordance index
      library (combinat)
      #memory.limit(size=5000)
      prog.c = combn(p.surls,2)
      time.c = combn(Time,2)
      delta.c = combn(delta,2)

      prog.c_1 = cbind(t(prog.c)[,1])
      prog.c_2 = cbind(t(prog.c)[,2])
      beda_prog.c = prog.c_2 - prog.c_1
    }
  }
}
```

**Lampiran 14.** Syntax untuk A-SURLSSVM dan Ukuran Performansinya (Lanjutan)

```
time.c_1 = cbind(t(time.c)[,1])
time.c_2 = cbind(t(time.c)[,2])
beda_time.c = time.c_2 - time.c_1

delta_ti = cbind(t(delta.c)[,1])
indikator = beda_prog.c*beda_time.c

for (k in 1:length(indikator))
{
  if (indikator[k]>0) {indikator[k]=1} else {indikator[k]=0}
}

comp = matrix (0, nrow=length(indikator), ncol=1)
for (l in 1:length(indikator))
{
  if (beda_time.c[l]>0 & delta_ti[l]==1)
  {comp[l]=1} else {comp[l]=0}
}

c.index = sum(indikator*comp)/sum(comp)*100
Perf.surls.i = cbind(mu = mu[mui], C = C[ci], kern_par = kern_par[ki],
                    HR = HR, Logrank = Logrank$chisq,
                    C.index = c.index)
Perf.surls = rbind(Perf.surls, Perf.surls.i)

Perbandinganx = data.frame(HR.SURLS = Perf.surls.i[4],
                           Logrank.SURLS = Perf.surls.i[5],
                           C.index.SURLS = Perf.surls.i[6],
                           mu = mu[mui], C = C[ci],
                           kern_par = kern_par[ki])
print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "D://bismillah tesis/hasil serviks/all
variabell.csv")

}
}
}
```

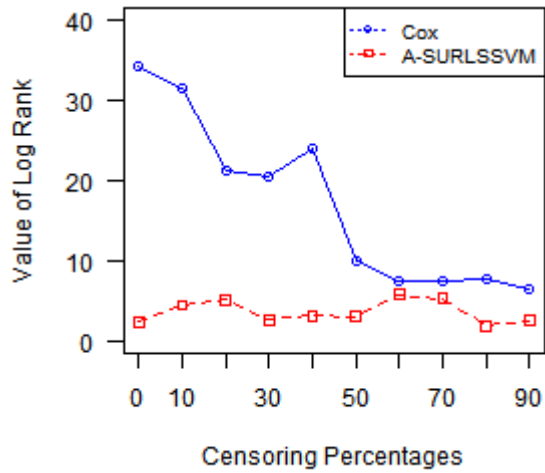
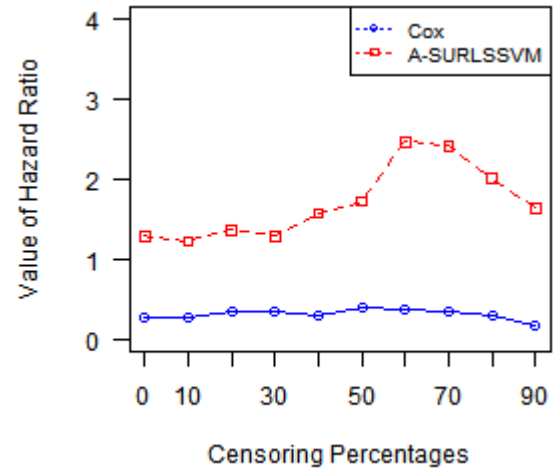
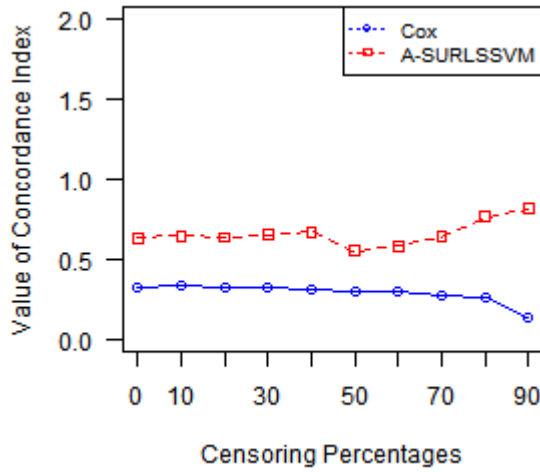


**Lampiran 15.** Syntax untuk Membuat Plot antara Performansi terhadap Persentase Sensoring

```
data = read.table("D:/data simulasi/censoring parametrik x=6 n=100.txt", header
= T)
data
Cox=data$cicox
ASURLSSVM=data$scisvm
g_range <- range(0:1)
maxy=max(Cox, ASURLSSVM)
# Graph autos using y axis that ranges from 0 to max
png('ci parametrik x=6 n=100.png',, height=305, width=300,bg="white")
plot(Cox, type="o", col="blue", ylim=g_range, axes=FALSE, ann=FALSE)
axis(1, at=1:9, lab=c("10","20","30","40","50","60","70","80","90"))
axis(2, las=1, at=NULL)
# Create box around plot
box()
# Graph trucks with red dashed line and square points
lines(ASURLSSVM, type="o", pch=22, lty=2, col="red")
# Create a title with a red, bold/italic font
#title(main="Autos", col.main="red", font.main=4)
# Label the x and y axes
title(xlab="Censoring Percentage", col.lab="black")
title(ylab="Value of C-Index", col.lab="black")
# Create a legend at (1, g_range[2]) that is slightly smaller
# (cex) and uses the same line colors and points used by
# the actual plots
legend("topright", c("Cox","A-SURLSSVM"), cex=0.8,
      col=c("blue","red"), pch=21:25, lty=3:4);
# Save graph
#png(filename="D:/censoringhrpar_100.png",height=305,
width=300,bg="white")
dev.off()
```

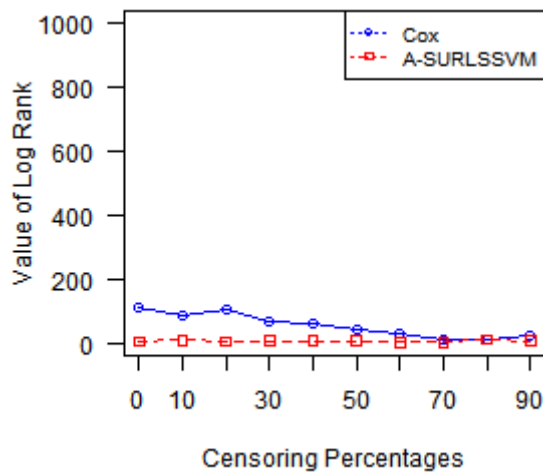
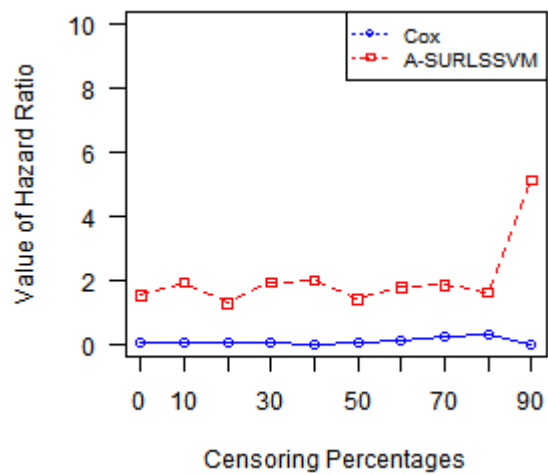
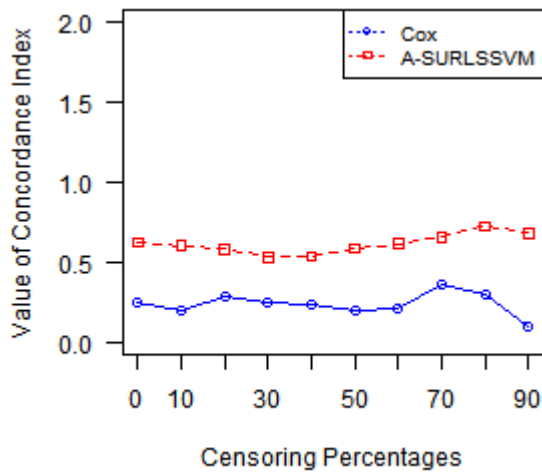
**Lampiran 16.** Performansi untuk Data Simulasi  $p=6$  dan  $n=100$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.274702	34.3	0.318791	1.283769	2.308406	0.63067
10	0.262464	31.4	0.329696	1.218266	4.419586	0.642789
20	0.348187	21.3	0.322309	1.360359	5.068538	0.63565
30	0.336558	20.4	0.324091	1.289894	2.621033	0.650031
40	0.289715	23.9	0.31321	1.569409	3.059979	0.667614
50	0.382007	9.9	0.291105	1.720523	3.001863	0.550764
60	0.377947	7.5	0.297685	2.464226	5.759876	0.579934
70	0.351679	7.5	0.276012	2.41073	5.22345	0.635838
80	0.285027	7.7	0.26177	2.001085	1.856981	0.764595
90	0.157188	6.4	0.130435	1.63229	2.541856	0.811594



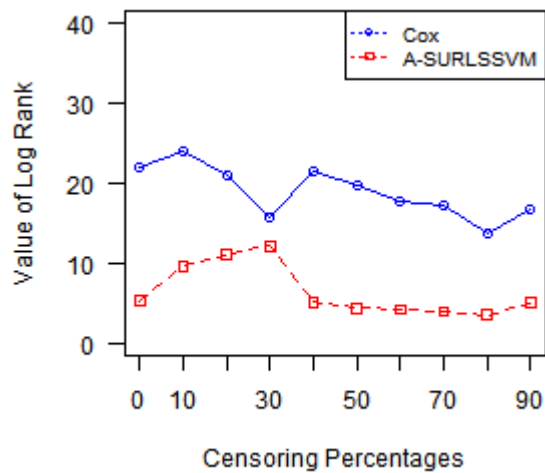
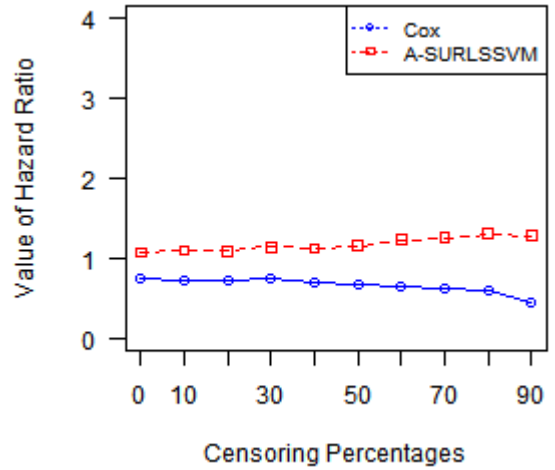
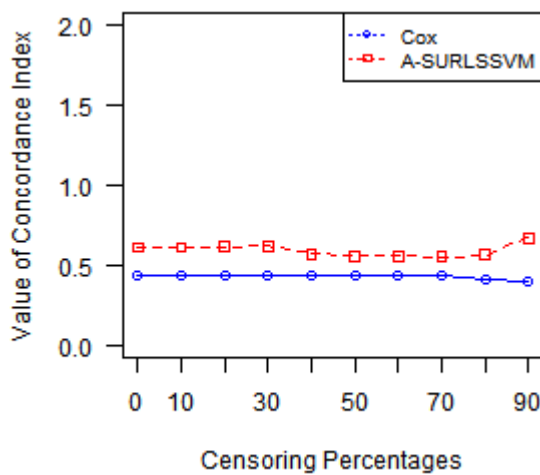
**Lampiran 17.** Performansi untuk Data Simulasi  $p=6$  dan  $n=100$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.034355	108	0.252232	1.533486	4.316285	0.621919
10	0.045005	85.8	0.197837	1.920344	8.443945	0.60132
20	0.010925	105	0.283902	1.275708	4.332315	0.578974
30	0.033407	66.2	0.249073	1.923402	7.107731	0.52892
40	0.008411	63.5	0.235591	1.988148	6.775171	0.53434
50	0.043493	40.4	0.200142	1.395233	5.615616	0.584084
60	0.132301	31.6	0.210223	1.778176	3.228444	0.613946
70	0.226345	10.8	0.358162	1.836831	2.45897	0.656
80	0.267838	10.9	0.299099	1.601765	10.08807	0.721101
90	1.98E-10	20.2	0.094059	5.106629	5.262151	0.678571



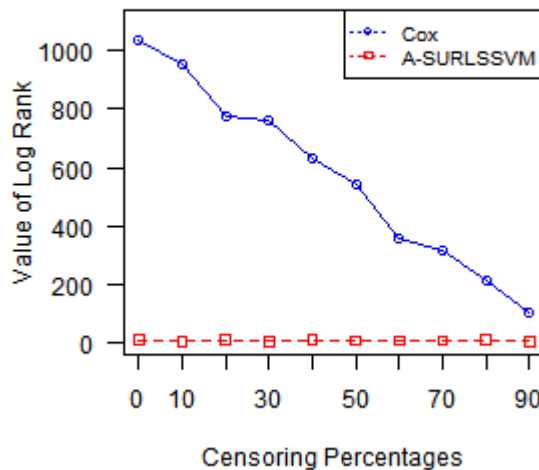
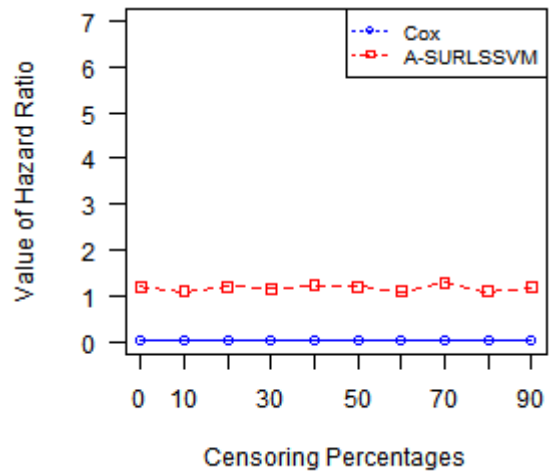
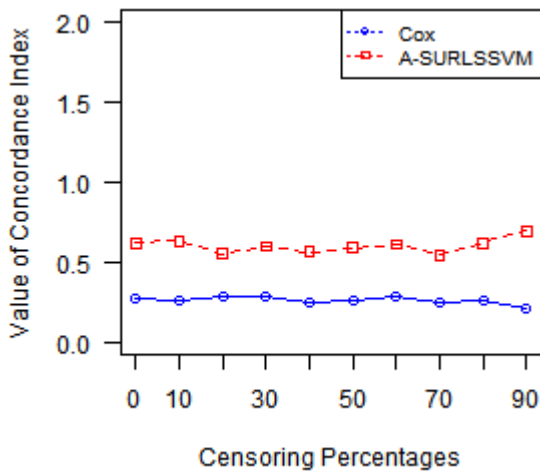
**Lampiran 18.** Performansi untuk Data Simulasi  $p=6$  dan  $n=1000$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.741934	21.9	0.434855	1.068058	5.215318	0.608724
10	0.71751	24	0.434861	1.093303	9.600831	0.610038
20	0.721017	21	0.435229	1.082452	11.02987	0.61147
30	0.737461	15.8	0.434038	1.132871	12.09556	0.618844
40	0.684196	21.4	0.428701	1.118605	5.111046	0.567416
50	0.672224	19.6	0.431995	1.147579	4.35352	0.554457
60	0.651756	17.7	0.4349	1.229189	4.19815	0.555248
70	0.620921	17.1	0.428579	1.253694	3.91931	0.548383
80	0.59454	13.6	0.414598	1.299973	3.480501	0.563004
90	0.446949	16.8	0.399666	1.274291	5.024345	0.668877



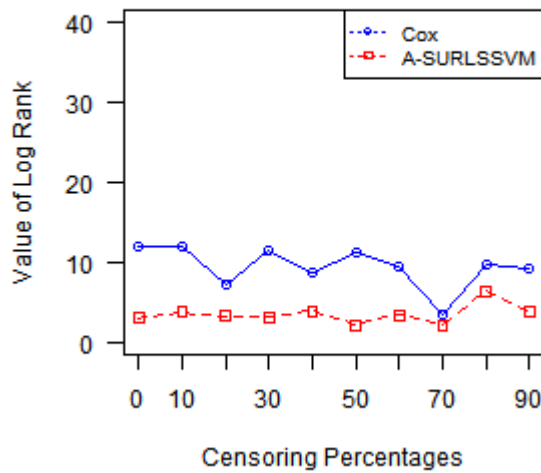
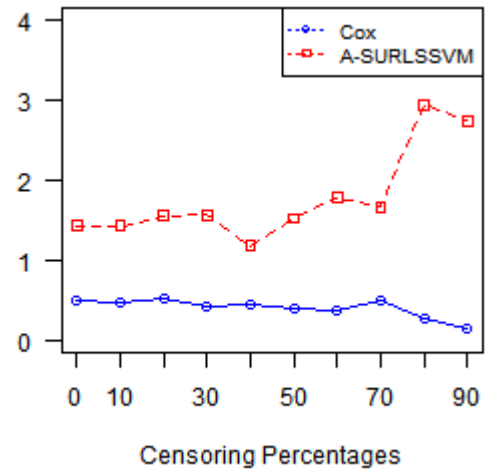
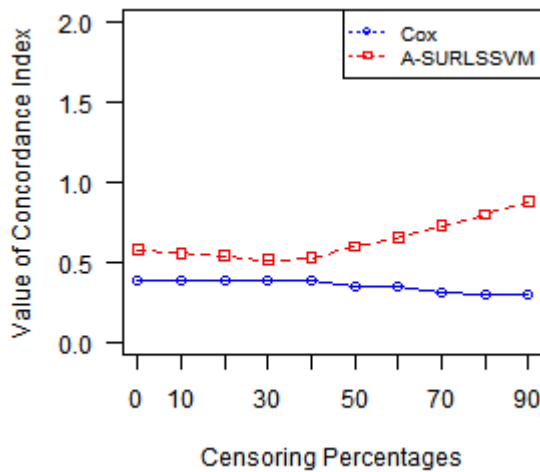
**Lampiran 19.** Performansi untuk Data Simulasi  $p=6$  dan  $n=1000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.039419	1039	0.275248	1.192377	7.573481	0.616991
10	0.034401	952	0.261612	1.089136	1.538718	0.631905
20	0.038676	773	0.282022	1.193763	6.377708	0.552295
30	0.032937	760	0.281062	1.150392	3.545935	0.597283
40	0.035555	632	0.249866	1.219019	5.724577	0.56415
50	0.03948	539	0.255485	1.194821	3.968448	0.586622
60	0.050551	359	0.281543	1.101716	4.819312	0.608696
70	0.039389	315	0.248993	1.281992	4.8679	0.544316
80	0.038101	210	0.25615	1.086719	7.714588	0.616991
90	0.045557	102	0.204567	1.177358	3.436044	0.69208



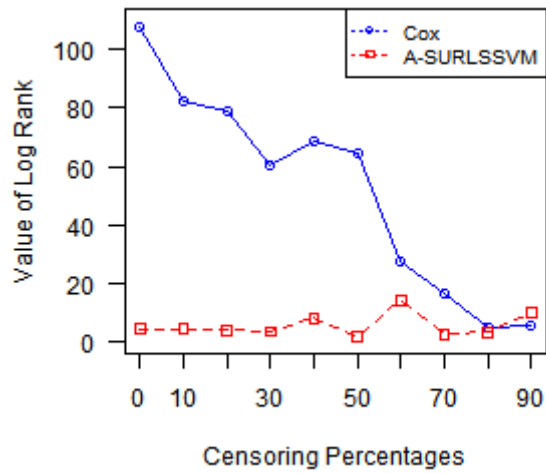
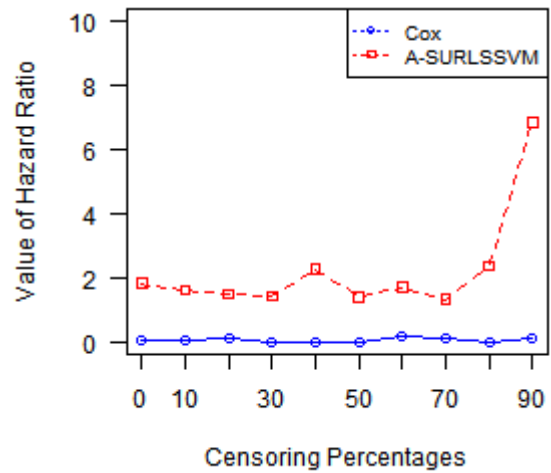
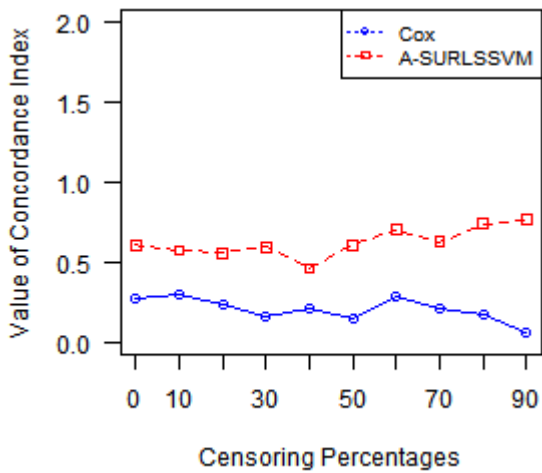
**Lampiran 20.** Performansi untuk Data Simulasi  $p=8$  dan  $n=100$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.48516	12	0.389798	1.422675	2.97762	0.576713
10	0.459056	11.9	0.379443	1.422442	3.724147	0.557637
20	0.522874	7.2	0.38914	1.551353	3.235176	0.536199
30	0.417004	11.4	0.379439	1.559197	3.067865	0.514019
40	0.438393	8.8	0.385144	1.174319	3.876193	0.52751
50	0.385416	11.1	0.345314	1.517025	2.086804	0.596899
60	0.378975	9.4	0.345601	1.77748	3.337819	0.64991
70	0.502062	3.5	0.311527	1.661666	2.099	0.727934
80	0.269495	9.8	0.29125	2.934035	6.360656	0.79625
90	0.138444	9.1	0.298354	2.742958	3.853078	0.876543



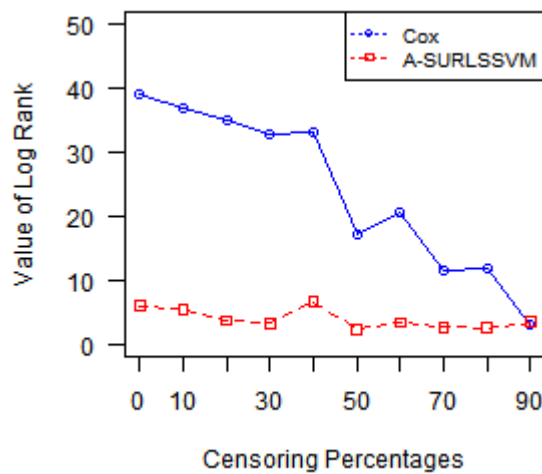
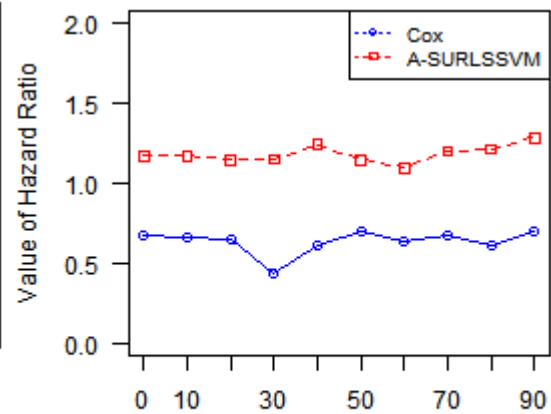
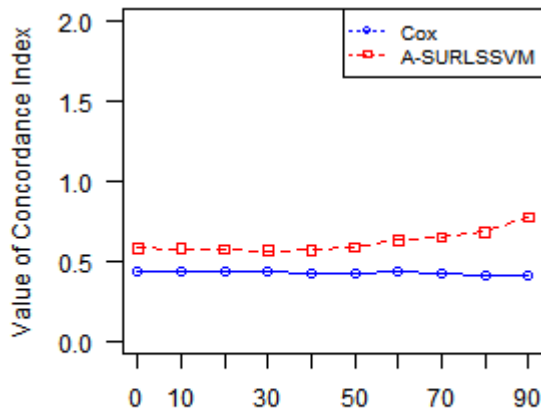
**Lampiran 21.** Performansi untuk Data Simulasi  $p=8$  dan  $n=100$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.016998	108	0.275578	1.804717	4.035062	0.607258
10	0.049741	82.1	0.295205	1.600835	4.013683	0.572944
20	0.076496	78.7	0.239142	1.478679	3.812726	0.557281
30	0.009757	60.4	0.160738	1.424331	3.270491	0.593952
40	0.010523	68.9	0.203151	2.248116	7.773776	0.459459
50	0.00799	64.3	0.15051	1.389374	1.225871	0.605521
60	0.137733	27.1	0.279963	1.700699	13.97398	0.698925
70	0.079572	16.4	0.213495	1.311979	2.269213	0.62624
80	3.56E-09	4.5	0.175743	2.363466	3.049541	0.736434
90	1.14E-01	5.4	0.063559	6.829041	9.640892	0.765217



**Lampiran 22.** Performansi untuk Data Simulasi  $p=8$  dan  $n=1000$  pada Data dengan Pola Linier

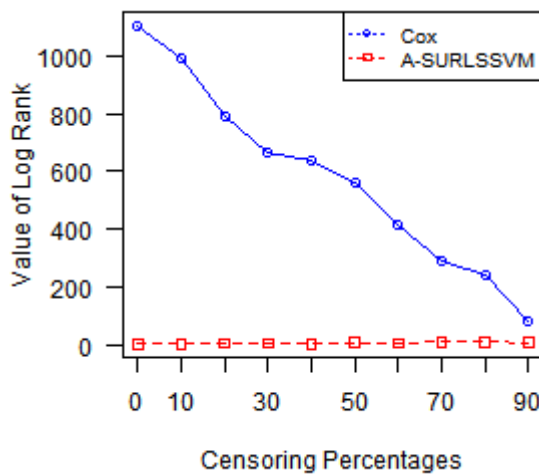
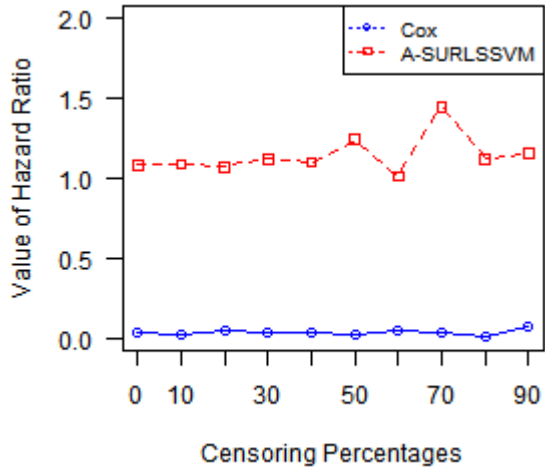
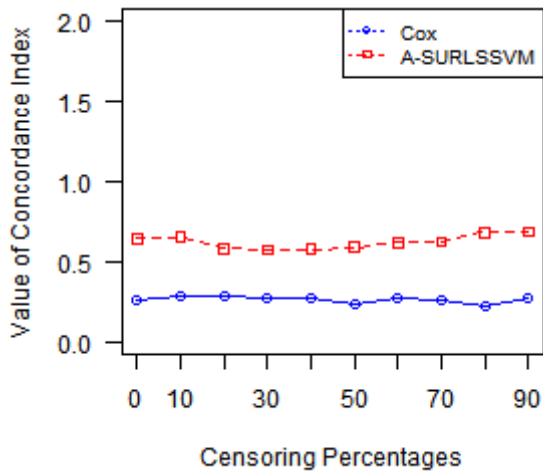
Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.666366	39	0.437349	1.170976	5.996331	0.579501
10	0.660005	36.9	0.436549	1.169631	5.4207	0.573292
20	0.650487	34.9	0.433515	1.145012	3.651222	0.572074
30	0.43137	32.7	0.43137	1.14877	3.224454	0.563497
40	0.611731	33.1	0.426219	1.240072	6.622013	0.567315
50	0.691358	17.2	0.424983	1.14252	2.265809	0.583419
60	0.637863	20.6	0.428682	1.093583	3.370343	0.630473
70	0.677769	11.4	0.423031	1.197986	2.575012	0.647543
80	0.605657	11.8	0.403542	1.210121	2.556901	0.679857
90	0.699572	3	0.406846	1.282682	3.385671	0.771562





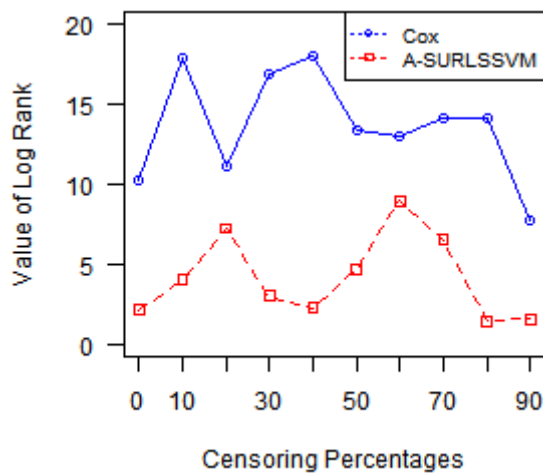
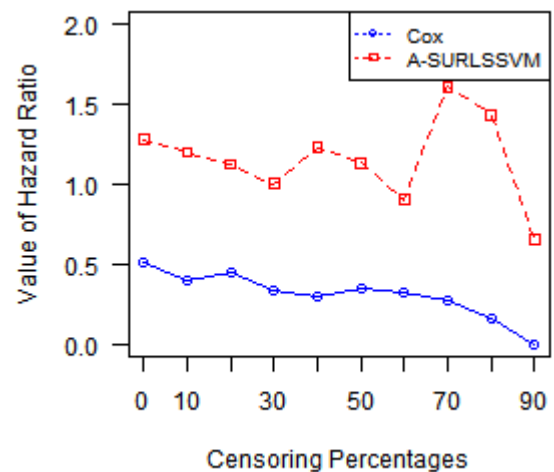
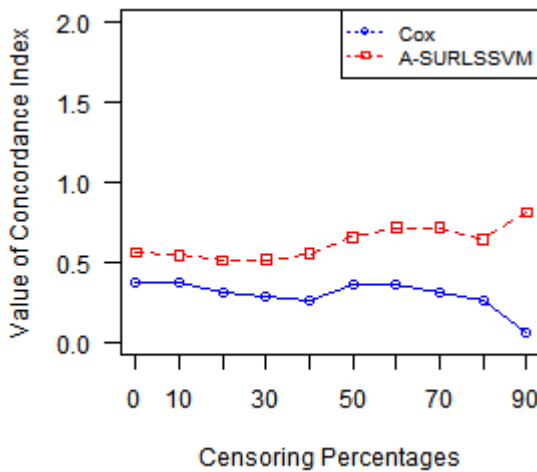
**Lampiran 23.** Performansi untuk Data Simulasi  $p=8$  dan  $n=1000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.029873	1102	0.263226	1.081043	3.739259	0.643184
10	0.024828	988	0.278457	1.083833	2.537987	0.648308
20	0.040104	788	0.288145	1.071148	4.947173	0.581769
30	0.032874	666	0.2768	1.119639	4.387389	0.570629
40	0.028819	636	0.275824	1.097223	3.751131	0.573347
50	0.025932	559	0.238107	1.2401	5.518947	0.587187
60	0.039974	416	0.27067	1.0113	4.856307	0.619244
70	0.03434	287	0.253334	1.44351	10.25817	0.621625
80	0.01265	240	0.220336	1.117393	9.891344	0.681911
90	0.065529	80.2	0.27595	1.151257	7.860364	0.685663



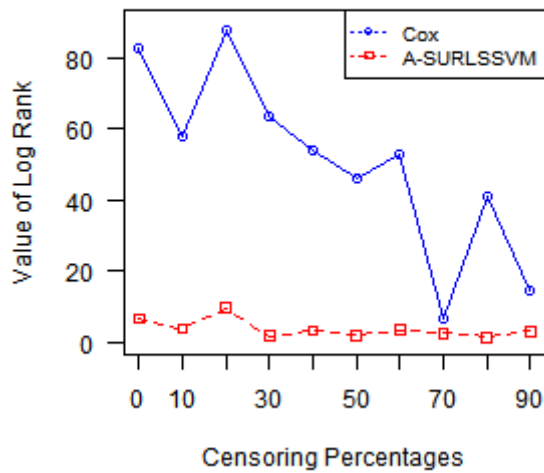
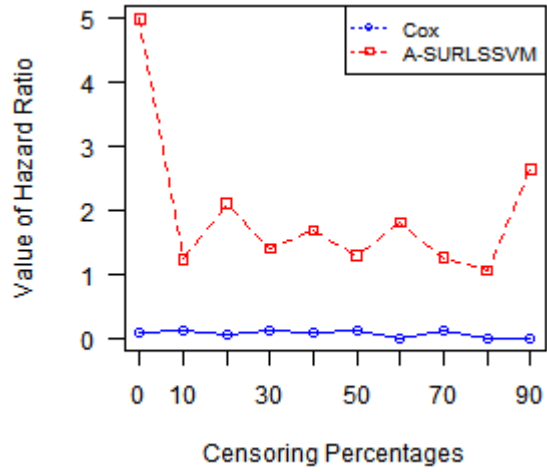
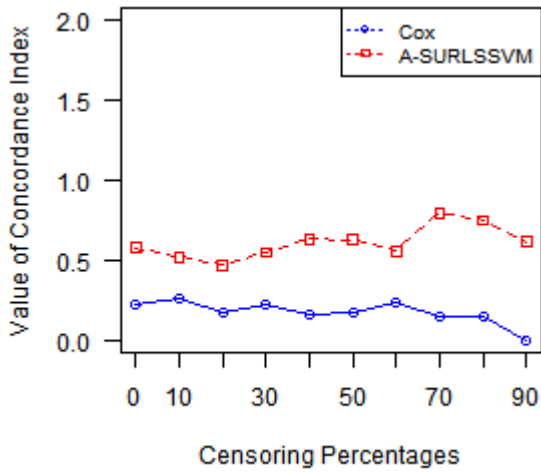
**Lampiran 24.** Performansi untuk Data Simulasi  $p=15$  dan  $n=100$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.514146	10.2	0.365338	1.277484	2.141936	0.560212
10	0.395849	17.9	0.368445	1.198779	4.008364	0.544151
20	0.450261	11.1	0.308977	1.122413	7.219369	0.509395
30	0.337672	16.8	0.285621	0.998875	2.984532	0.51346
40	0.299304	18	0.259797	1.228625	2.252166	0.552975
50	0.351921	13.4	0.355607	1.132884	4.667326	0.655914
60	0.320546	13	0.355556	0.900713	8.947846	0.712037
70	0.26782	14.1	0.31123	1.609439	6.499603	0.715508
80	0.16318	14.1	0.252768	1.430201	1.435099	0.643911
90	1.54E-09	7.7	0.061674	0.653741	1.568308	0.810573



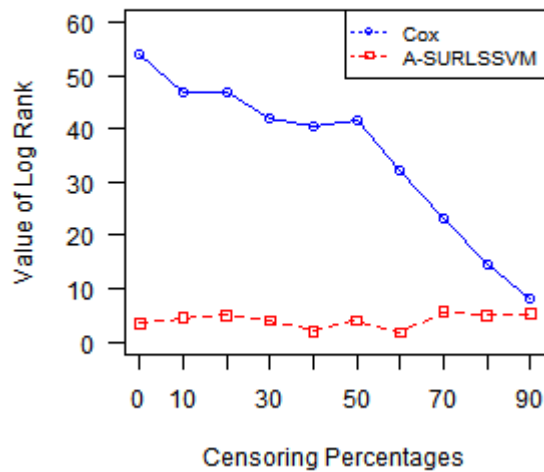
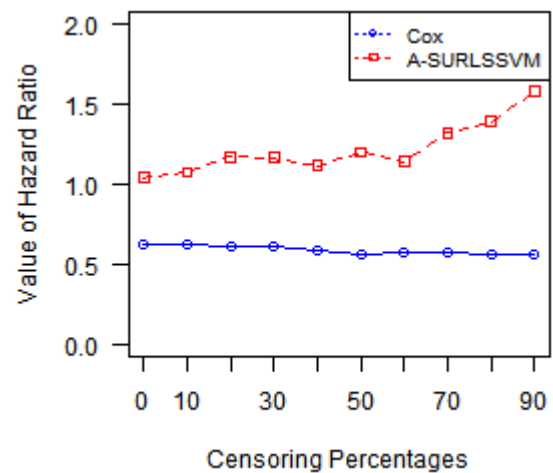
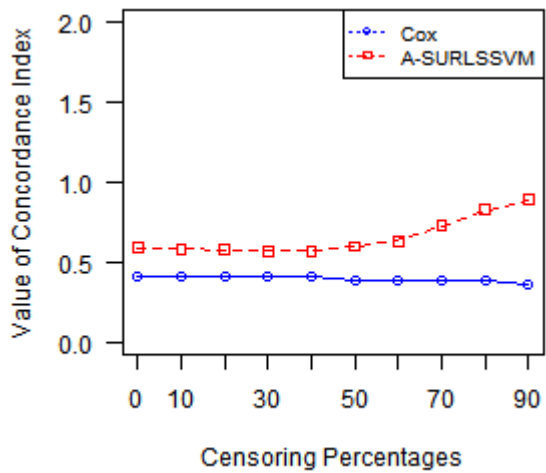
**Lampiran 25.** Performansi untuk Data Simulasi  $p=15$  dan  $n=100$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.075905	82.5	0.221658	4.987712	6.54157	0.574522
10	0.128644	58	0.256206	1.231289	3.701965	0.520352
20	0.037526	87.7	0.170455	2.101574	9.605715	0.467637
30	0.104506	63.2	0.224713	1.395641	1.732392	0.546978
40	0.086125	54	0.155512	1.680634	3.2488	0.63174
50	0.125536	45.8	0.174222	1.285084	1.828504	0.626962
60	2.55E-10	52.8	0.232955	1.806129	3.29071	0.557673
70	0.109287	6.7	0.143434	1.249157	2.562203	0.792373
80	7.47E-10	40.9	0.141089	1.055296	1.271991	0.746442
90	5.67E-10	14.6	0	2.631582	3.073303	0.615023



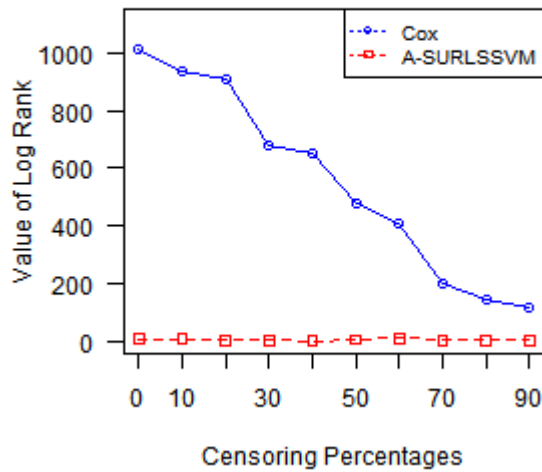
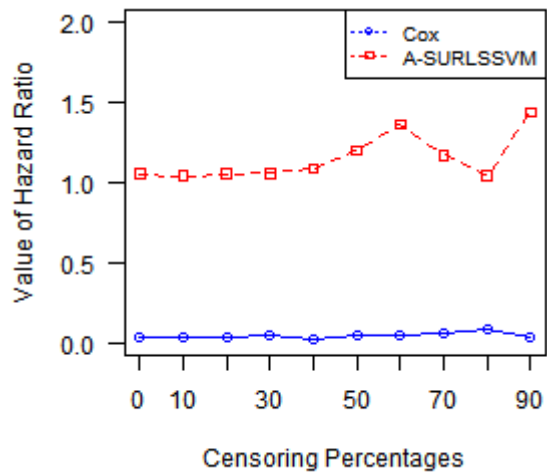
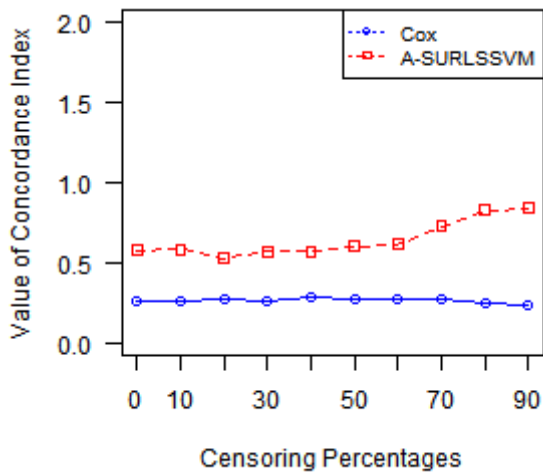
**Lampiran 26.** Performansi untuk Data Simulasi  $p=15$  dan  $n=1000$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.623568	54.1	0.410548	1.039885	3.549376	0.588756
10	0.627497	46.9	0.40855	1.073712	4.568038	0.580524
20	0.613642	46.7	0.409594	1.171165	5.056236	0.573329
30	0.609794	41.9	0.408851	1.166026	4.028876	0.566266
40	0.586176	40.3	0.406534	1.113655	2.024384	0.570307
50	0.556427	41.6	0.385881	1.197066	4.007615	0.597759
60	0.568486	32.3	0.379139	1.141138	1.772715	0.628378
70	0.56578	23.3	0.383379	1.320445	5.640246	0.725963
80	0.565011	14.7	0.384982	1.390776	5.060255	0.827503
90	0.553432	8.2	0.353402	1.576636	5.321405	0.887202



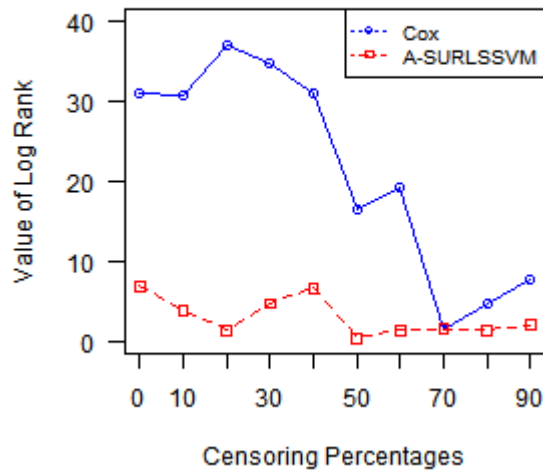
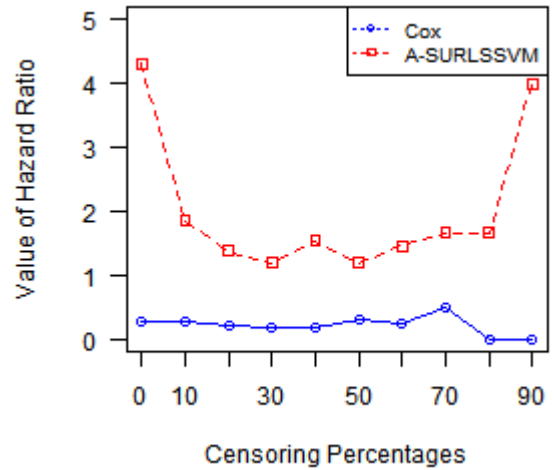
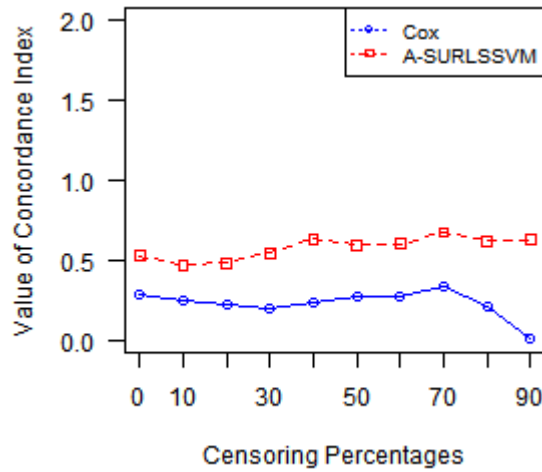
**Lampiran 27.** Performansi untuk Data Simulasi  $p=15$  dan  $n=1000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.036518	1009	0.259002	1.053013	7.475143	0.576966
10	0.034313	939	0.256428	1.038296	5.431488	0.579168
20	0.030731	910	0.273524	1.053339	2.138061	0.524557
30	0.041034	676	0.260559	1.054382	2.738141	0.568205
40	0.025899	648	0.285119	1.084885	0.977209	0.566163
50	0.040675	480	0.268889	1.199024	4.119543	0.599227
60	0.041392	406	0.267267	1.361222	9.088697	0.612114
70	0.06057	202	0.273319	1.167368	2.917134	0.72681
80	0.084758	145	0.240797	1.038975	3.065777	0.824486
90	0.029932	118	0.237137	1.43465	3.529274	0.837675



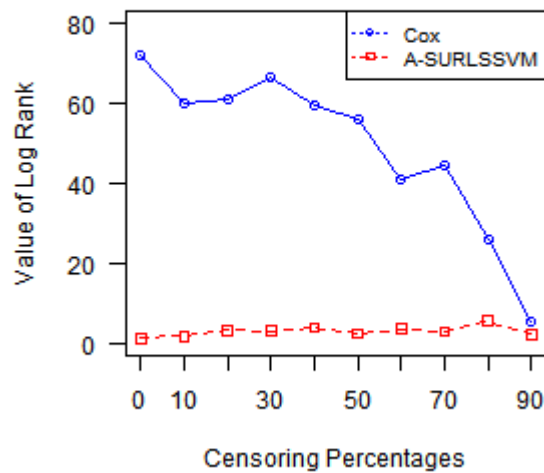
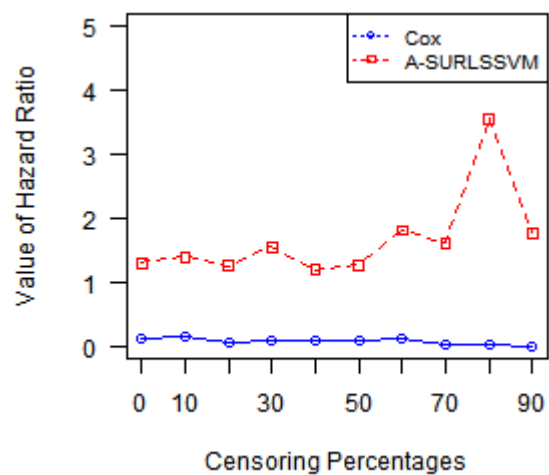
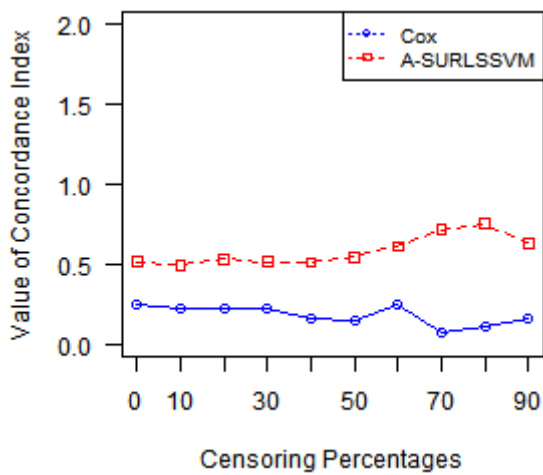
**Lampiran 28.** Performansi untuk Data Simulasi  $p=17$  dan  $n=100$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.282367	31	0.278268	4.302134	6.816293	0.527022
10	0.270055	30.8	0.243255	1.841868	3.875139	0.467798
20	0.197414	36.9	0.215228	1.373248	1.38326	0.481726
30	0.181485	34.6	0.198881	1.184947	4.652898	0.545059
40	0.17861	30.9	0.230116	1.529352	6.636993	0.630116
50	0.302308	16.5	0.270184	1.182511	0.380488	0.590591
60	0.235248	19.1	0.272451	1.446562	1.389416	0.601979
70	0.480539	1.5	0.34	1.662585	1.542888	0.672
80	4.03E-09	4.7	0.203443	1.662574	1.37627	0.616588
90	2.09E-09	7.8	0.003215	3.991495	2.016687	0.62701



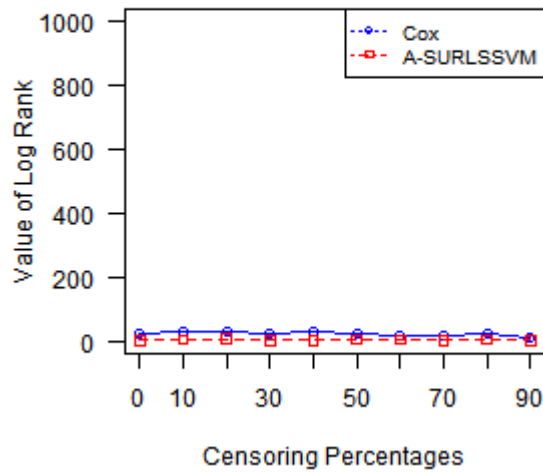
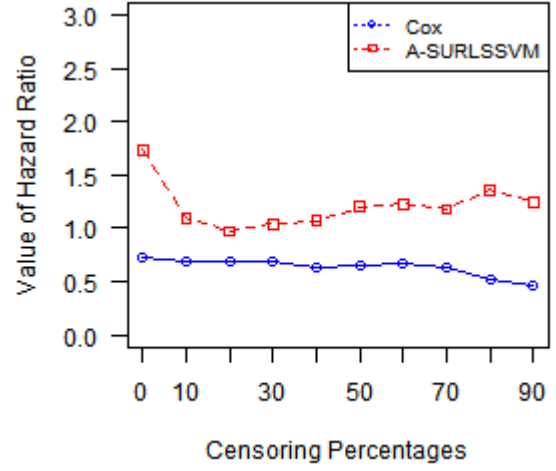
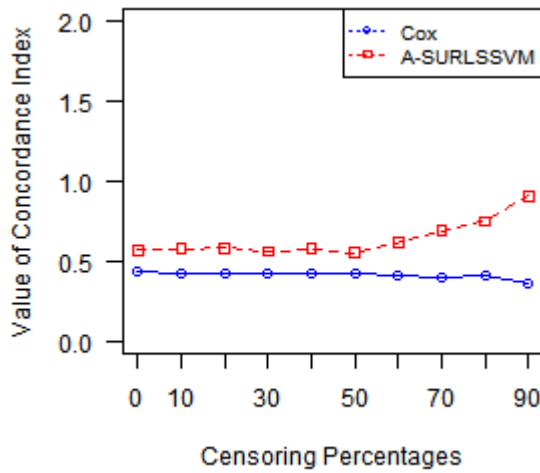
**Lampiran 29.** Performansi untuk Data Simulasi  $p=17$  dan  $n=100$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.105767	71.9	0.25146	1.29101	1.271155	0.512495
10	0.14955	59.9	0.224455	1.386637	1.591314	0.491899
20	0.066257	60.8	0.22095	1.256309	3.139449	0.532722
30	0.082714	66.4	0.219209	1.545309	2.893936	0.51398
40	0.089728	59.6	0.152893	1.187512	3.833621	0.509091
50	0.083745	56.1	0.14472	1.265719	2.29483	0.544832
60	1.07E-01	41.1	0.244385	1.802777	3.479616	0.610287
70	0.029943	44.3	0.073583	1.61049	2.796479	0.712021
80	2.88E-02	25.7	0.108808	3.552311	5.400314	0.75
90	3.11E-09	5.1	0.154054	1.76762	2.132842	0.632768



**Lampiran 30.** Performansi untuk Data Simulasi  $p=17$  dan  $n=1000$  pada Data dengan Pola Linier

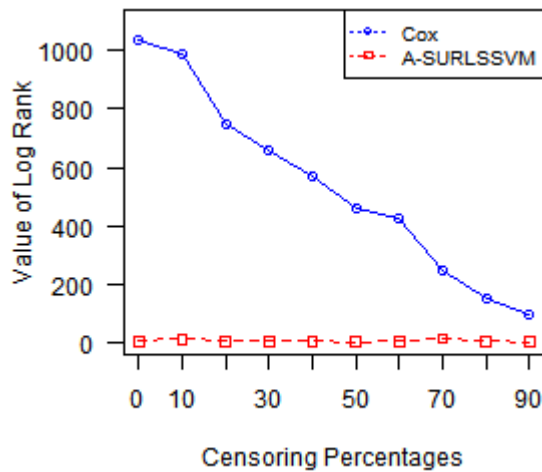
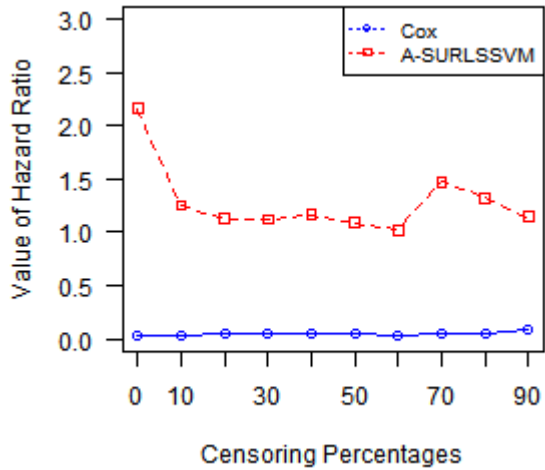
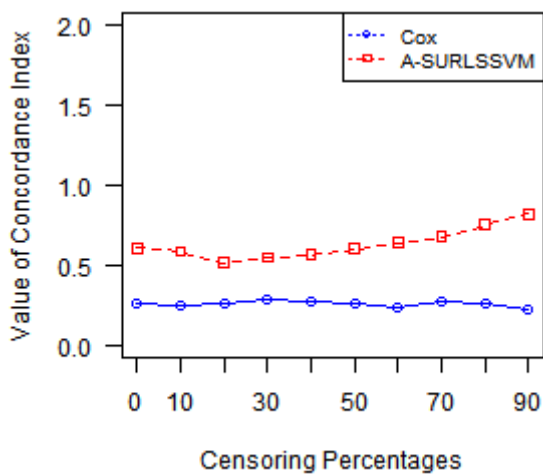
Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.730765	24.3	0.431103	1.736682	1.356759	0.569153
10	0.69189	30.2	0.423711	1.09616	4.032739	0.577297
20	0.692995	26.8	0.420649	0.972335	6.874087	0.57971
30	0.698008	22.4	0.426341	1.0381	3.370727	0.560333
40	0.633768	29.7	0.423899	1.06683	2.897863	0.575123
50	0.658671	22.8	0.425297	1.203772	4.512293	0.552497
60	0.676774	16.1	0.414059	1.229206	4.507115	0.613838
70	0.629857	16.3	0.401033	1.17604	1.987031	0.688578
80	0.519737	20.5	0.408129	1.359941	4.572308	0.747442
90	0.470466	12.8	0.354897	1.246564	1.16418	0.906255





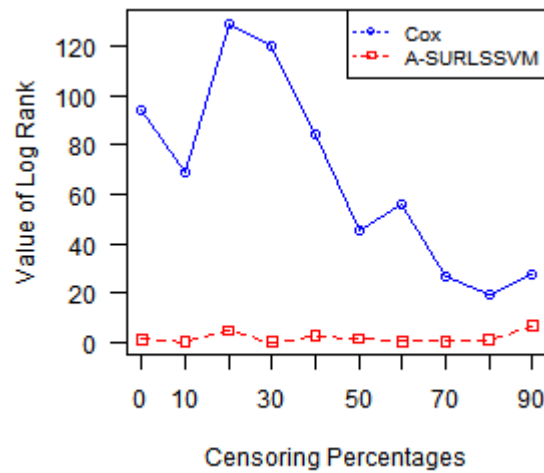
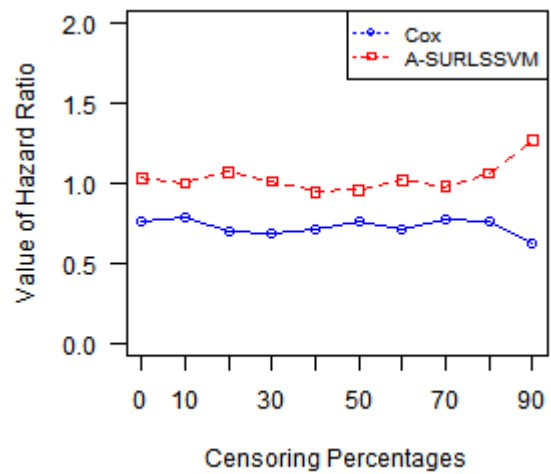
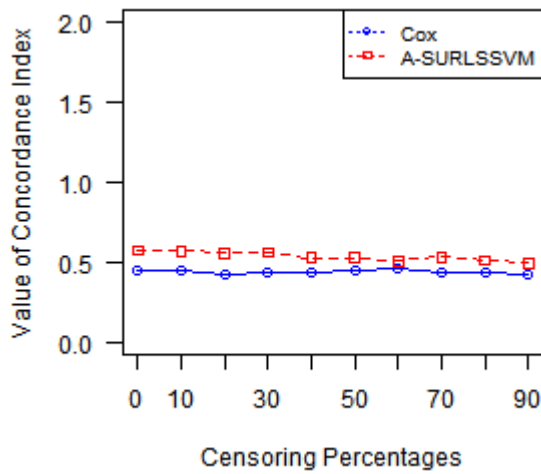
**Lampiran 31.** Performansi untuk Data Simulasi  $p=17$  dan  $n=1000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.033314	1036	0.262894	2.158896	3.095976	0.605949
10	0.027218	986	0.242702	1.250178	8.350965	0.584371
20	0.046431	746	0.261834	1.122739	2.570884	0.515032
30	0.04081	656	0.284539	1.120909	2.262905	0.545971
40	0.045097	567	0.274363	1.163536	3.48696	0.560886
50	0.047124	461	0.263787	1.091	1.002638	0.598465
60	0.032427	424	0.238874	1.020989	2.903438	0.637665
70	0.056087	245	0.266575	1.470348	11.10368	0.676186
80	0.058068	150	0.262399	1.322999	3.512627	0.75242
90	0.094678	93.3	0.225778	1.141448	0.490679	0.819968



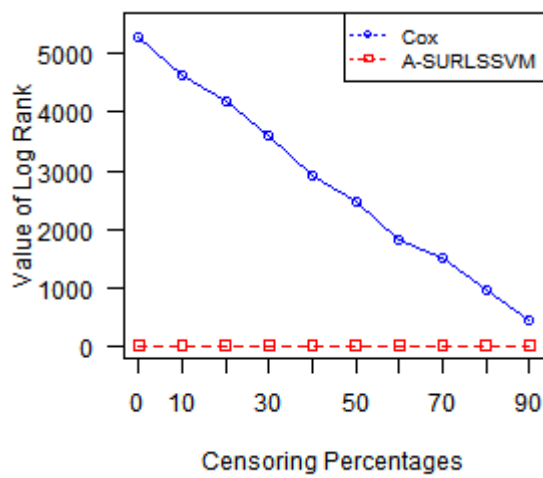
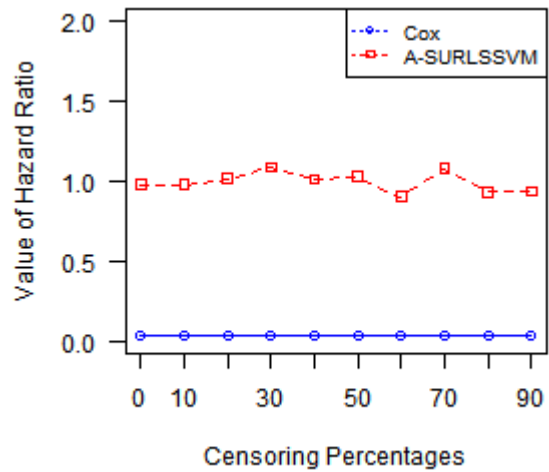
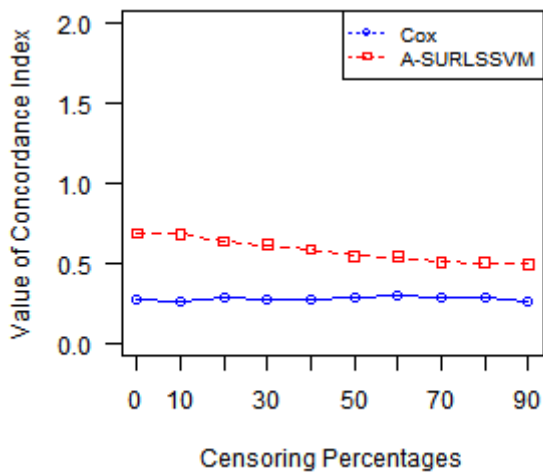
**Lampiran 32.** Performansi untuk Data Simulasi  $p=6$  dan  $n=5000$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.758629	94	0.449227	1.030343	1.112731	0.572651
10	0.780691	68.6	0.446467	0.998568	0.002284	0.566057
20	0.696896	129	0.427637	1.070665	4.601582	0.553328
30	0.688403	120	0.439053	1.010265	0.091796	0.559267
40	0.709885	84.7	0.434909	0.944855	2.35296	0.525761
50	0.764784	45.7	0.450385	0.953882	1.437785	0.52684
60	0.713965	55.9	0.457403	1.017481	0.151766	0.506152
70	0.767374	26.5	0.438235	0.977371	0.199862	0.532767
80	0.754911	19.5	0.43808	1.060073	0.840679	0.508983
90	0.623936	27.2	0.419863	1.265321	6.45845	0.495091



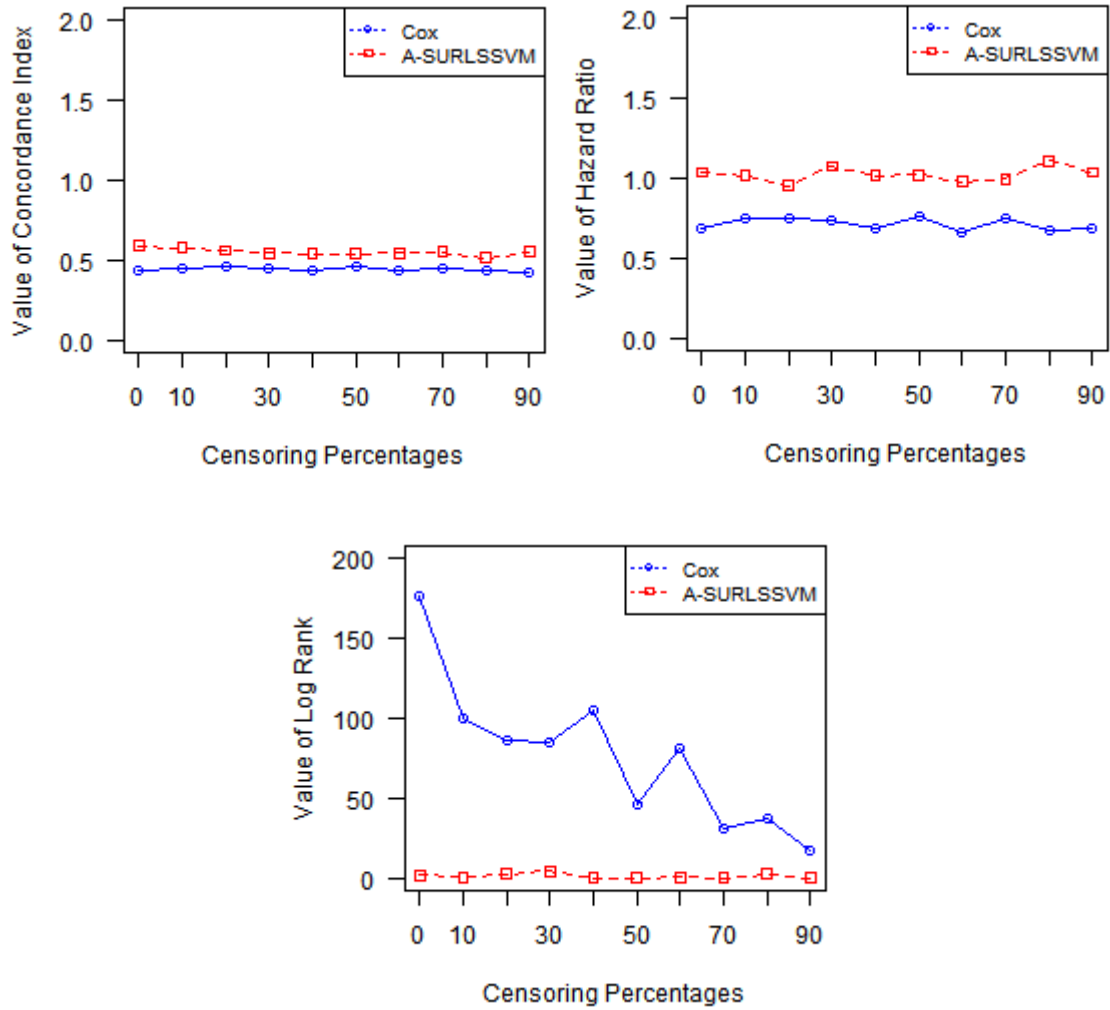
**Lampiran 33.** Performansi untuk Data Simulasi  $p=6$  dan  $n=5000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.034334	5292	0.27621	0.977051	0.669697	0.685372
10	0.032609	4645	0.263661	0.974291	0.753368	0.679051
20	0.03303	4174	0.282236	1.013556	0.174214	0.63533
30	0.036524	3617	0.267438	1.085325	5.797371	0.612404
40	0.036305	2928	0.270659	1.010952	0.087545	0.583424
50	0.037544	2482	0.280573	1.028703	0.490754	0.543255
60	0.039632	1799	0.292061	0.899451	5.144935	0.535662
70	0.036753	1518	0.280645	1.075985	1.979694	0.505382
80	0.038958	944	0.279043	0.929263	1.27228	0.498403
90	0.030905	445	0.260502	0.9361	0.508622	0.491788



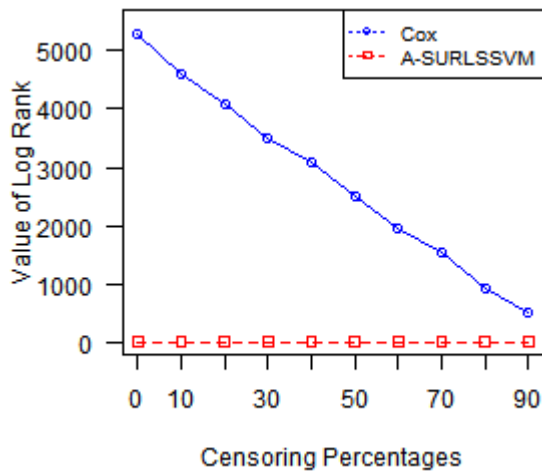
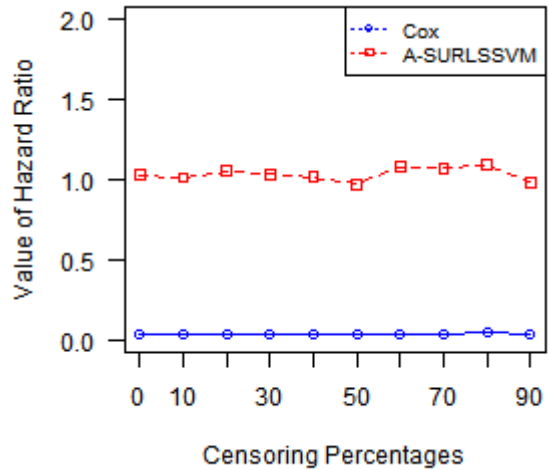
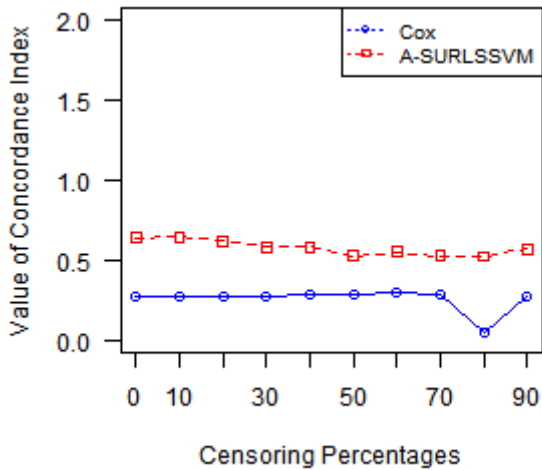
**Lampiran 34.** Performansi untuk Data Simulasi  $p=8$  dan  $n=5000$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.68459	176	0.435917	1.035859	1.548179	0.585767
10	0.741485	99.8	0.449681	1.014248	0.225786	0.577329
20	0.743473	85.8	0.457019	0.950079	2.610086	0.558793
30	0.730262	84.3	0.441214	1.072647	4.255006	0.544924
40	0.683437	105	0.428796	1.012654	0.116545	0.539813
50	0.761372	45.3	0.454146	1.018116	0.198101	0.540319
60	0.662828	81.2	0.436811	0.976488	0.278635	0.542225
70	0.746909	31.3	0.443737	0.992608	0.020463	0.549909
80	0.674323	37.2	0.4382	1.105141	2.456609	0.515395
90	0.684257	17.3	0.415421	1.035421	0.14571	0.552547



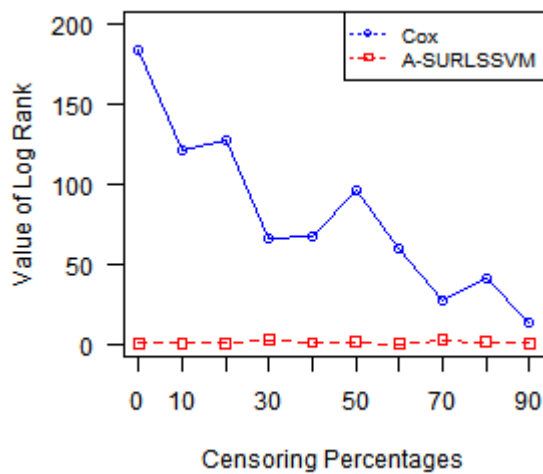
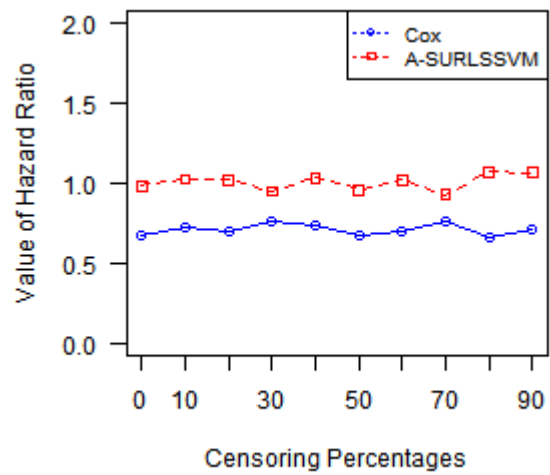
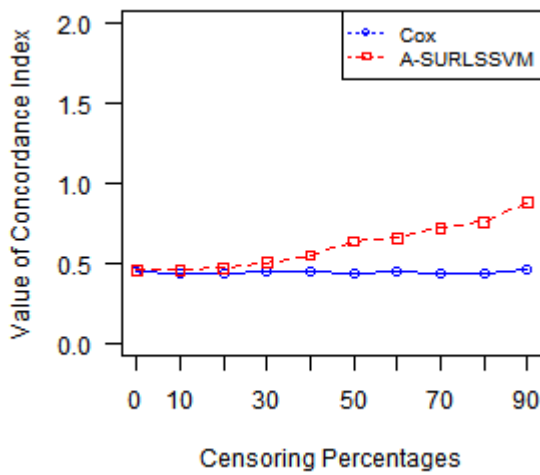
**Lampiran 35.** Performansi untuk Data Simulasi  $p=8$  dan  $n=5000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.033566	5300	0.272983	1.025746	0.783169	0.639969
10	0.036146	4597	0.269203	1.009256	0.092625	0.641714
20	0.035112	4085	0.274155	1.053741	2.5892	0.616691
30	0.038452	3512	0.274481	1.032189	0.867335	0.581775
40	0.035027	3092	0.279488	1.011918	0.106531	0.579672
50	0.037184	2496	0.279574	0.968622	0.62527	0.523786
60	0.038087	1951	0.292484	1.074396	2.568619	0.549381
70	0.037506	1526	0.279922	1.068663	1.714653	0.527621
80	0.046317	902	0.046317	1.091218	1.884713	0.520407
90	0.038296	506	0.268812	0.980404	0.047853	0.569689



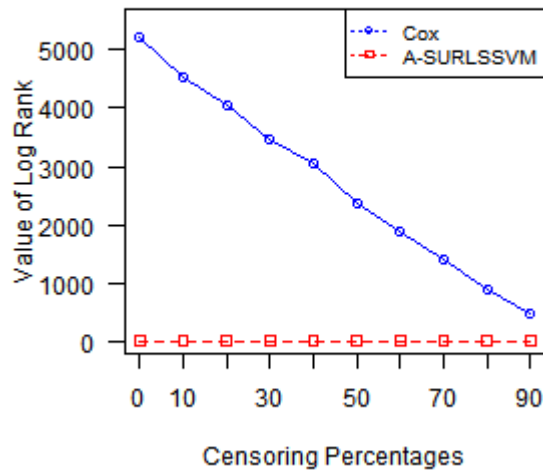
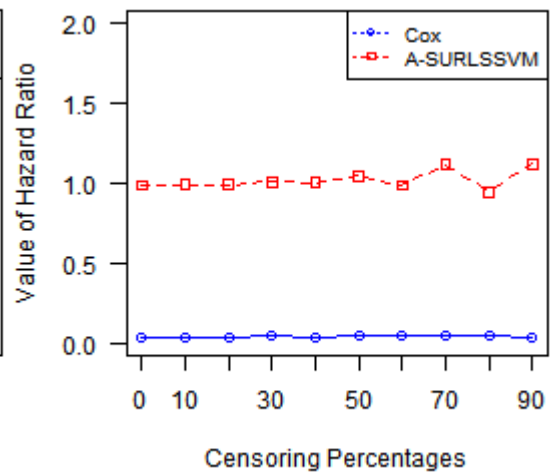
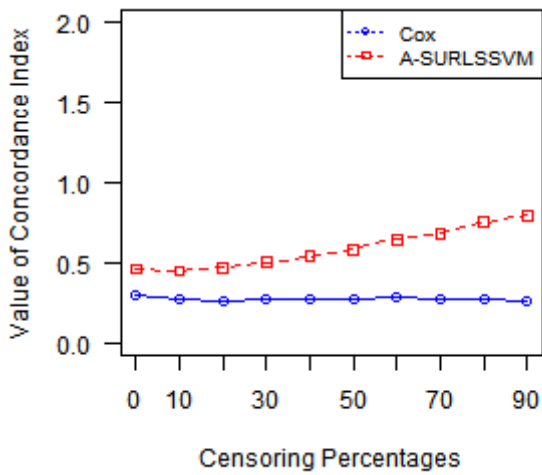
**Lampiran 36.** Performansi untuk Data Simulasi  $p=15$  dan  $n=5000$  pada Data dengan Pola Linier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.677159	184	0.442178	0.981699	0.417527	0.456762
10	0.718781	121	0.43916	1.023209	0.588662	0.453604
20	0.697779	127	0.433041	1.018183	0.324034	0.469596
30	0.758356	65.7	0.448431	0.94684	2.575872	0.502381
40	0.739111	66.7	0.442525	1.032077	0.733188	0.546351
50	0.673673	96.6	0.433581	0.953881	1.388288	0.63488
60	0.701218	60.3	0.440756	1.016393	0.125526	0.655915
70	0.76547	26.6	0.438574	0.924162	2.290019	0.715877
80	0.663068	41.4	0.439467	1.068753	1.086622	0.757667
90	0.713255	13.8	0.453531	1.065955	0.484622	0.875237



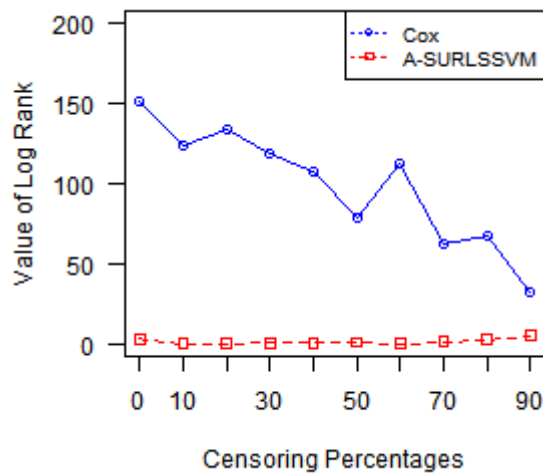
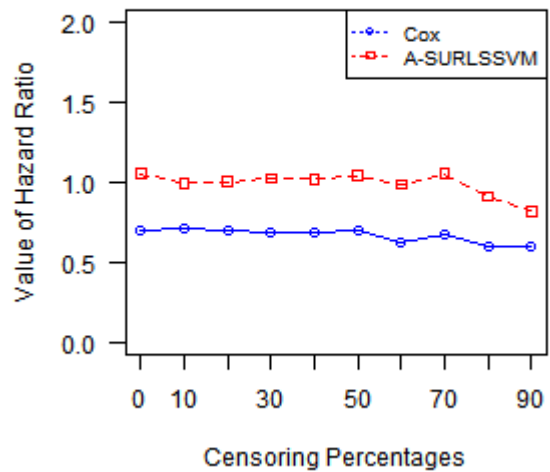
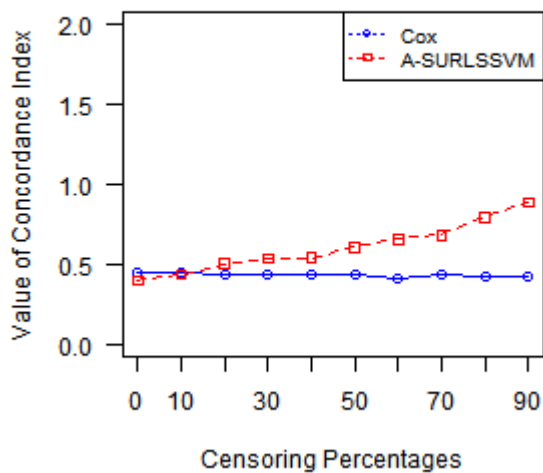
**Lampiran 37.** Performansi untuk Data Simulasi  $p=15$  dan  $n=5000$  pada Data dengan Pola Nonlinier

Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.033552	5216	0.290947	0.986183	0.244403	0.459532
10	0.037912	4545	0.272421	0.99066	0.08152	0.44651
20	0.036926	4068	0.260304	0.987902	0.148425	0.466809
30	0.041542	3472	0.273461	1.007795	0.067585	0.498186
40	0.036548	3057	0.272727	0.999048	0.00222	0.536798
50	0.041489	2355	0.269232	1.042085	0.995972	0.580463
60	0.040658	1876	0.283297	0.984039	0.138552	0.642551
70	0.042654	1409	0.268569	1.112809	4.323198	0.682438
80	0.049157	901	0.276542	0.944448	0.807939	0.752902
90	0.036679	473	0.263532	1.116543	1.432621	0.793107



**Lampiran 38.** Performansi untuk Data Simulasi  $p=17$  dan  $n=5000$  pada Data dengan Pola Linier

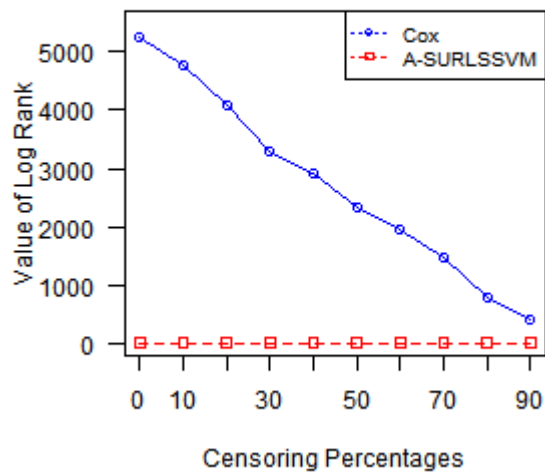
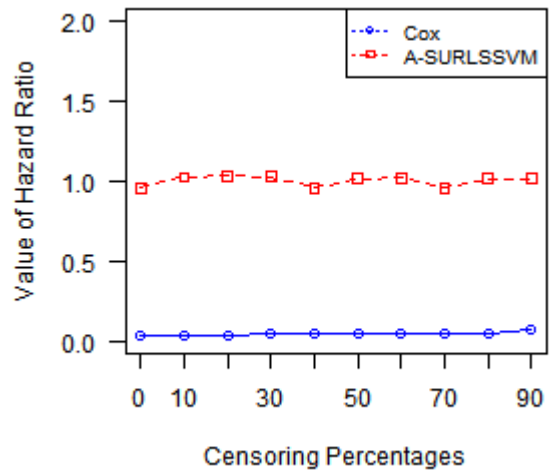
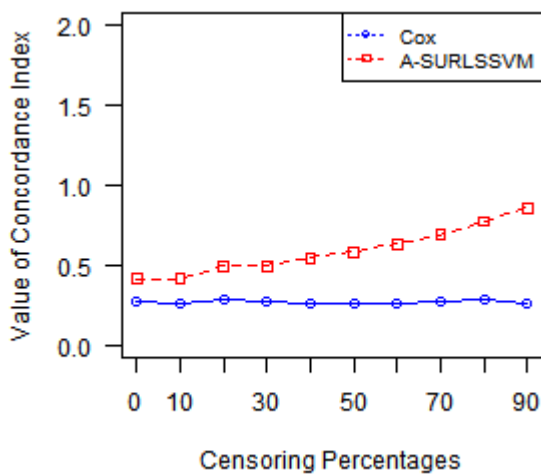
Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.703134	151	0.440523	1.050232	2.74892	0.396814
10	0.715306	123	0.440537	0.993336	0.046258	0.433107
20	0.691532	134	0.435737	1.001145	0.001637	0.500372
30	0.690204	118	0.437487	1.022564	0.427279	0.533609
40	0.683028	107	0.434063	1.020973	0.317727	0.53625
50	0.698911	78.9	0.429354	1.038503	0.890258	0.607762
60	0.615963	112	0.412716	0.986206	0.094503	0.656423
70	0.66704	62.2	0.432443	1.052394	0.987224	0.680791
80	0.598556	67	0.424562	0.909577	2.253328	0.792189
90	0.601811	32.3	0.418197	0.816378	4.805818	0.883656



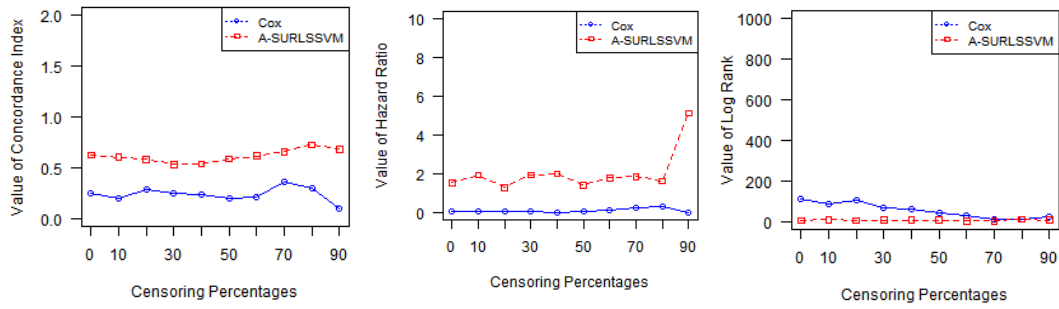


**Lampiran 39.** Performansi untuk Data Simulasi  $p=17$  dan  $n=5000$  pada Data dengan Pola Nonlinier

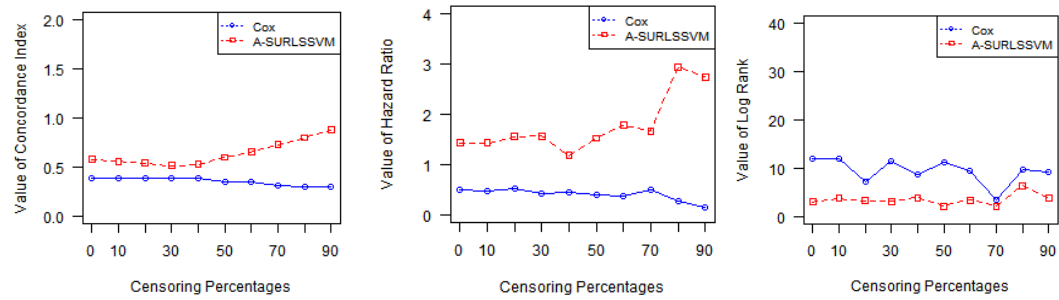
Persentase Tersensor	hrcox	lrcox	cicox	hrsvm	lrsvm	cisvm
0	0.033359	5253	0.273746	0.954365	2.643032	0.415261
10	0.030934	4778	0.257997	1.021421	0.476942	0.413957
20	0.036544	4098	0.281199	1.030615	0.942274	0.495063
30	0.04322	3283	0.271867	1.024312	0.467607	0.491307
40	0.04127	2927	0.260955	0.95466	1.696862	0.541776
50	0.041474	2328	0.261639	1.014892	0.139149	0.581354
60	0.039755	1968	0.264078	1.020655	0.225159	0.630645
70	0.042031	1486	0.266049	0.954967	0.833019	0.686803
80	0.048545	776	0.278548	1.015164	0.053148	0.770963
90	0.066827	417	0.257027	1.015818	0.030479	0.854896



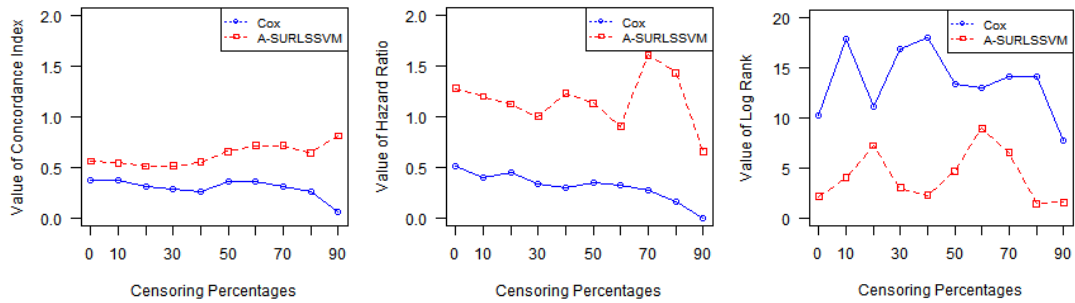
**Lampiran 40 a.** Performansi untuk Data Simulasi  $p=6$   $n=100$  dengan Pola Linier



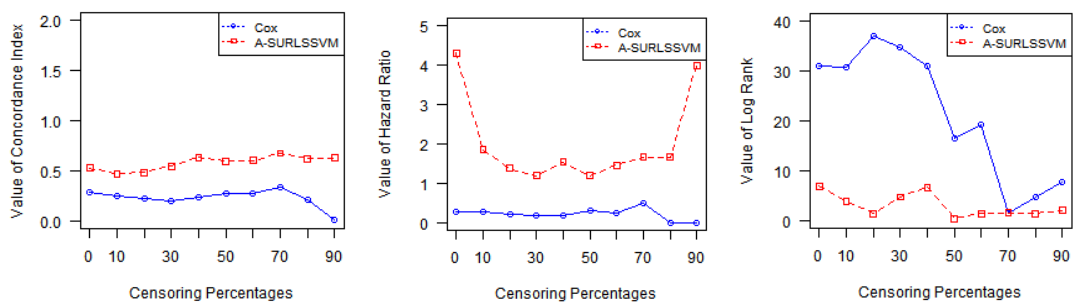
**Lampiran 40 b.** Performansi untuk Data Simulasi  $p=8$   $n=100$  dengan Pola Linier



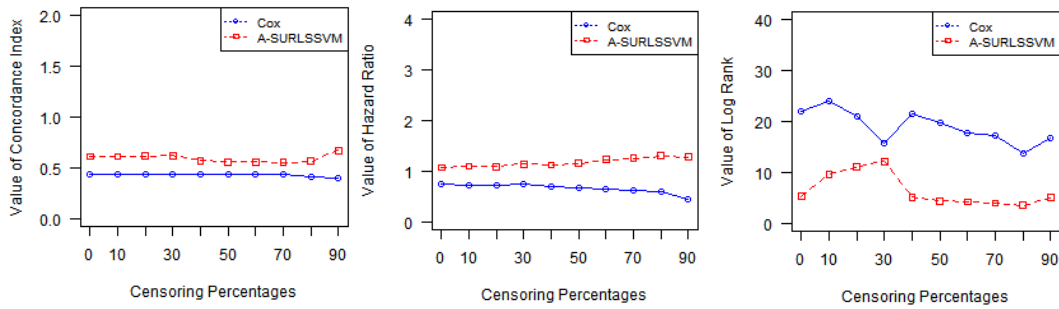
**Lampiran 40 c.** Performansi untuk Data Simulasi  $p=15$   $n=100$  dengan Pola Linier



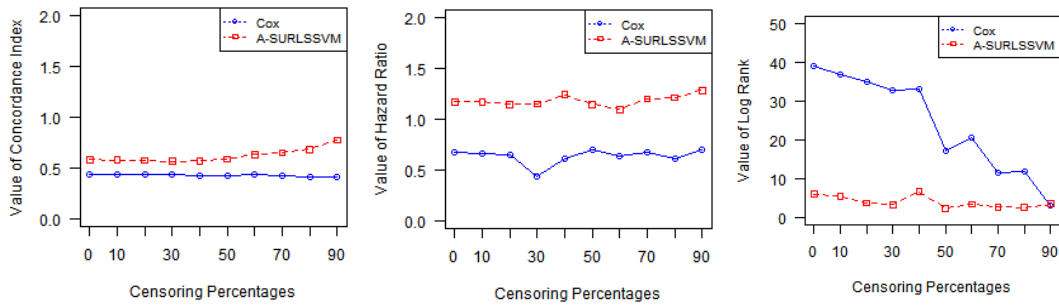
**Lampiran 40 d.** Performansi untuk Data Simulasi  $p=17$   $n=100$  dengan Pola Linier



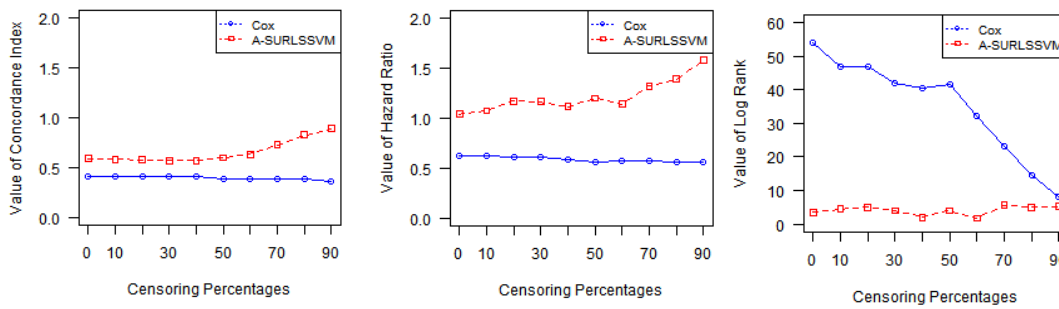
**Lampiran 41 a.** Performansi untuk Data Simulasi  $p=6$   $n=1000$  dengan Pola Linier



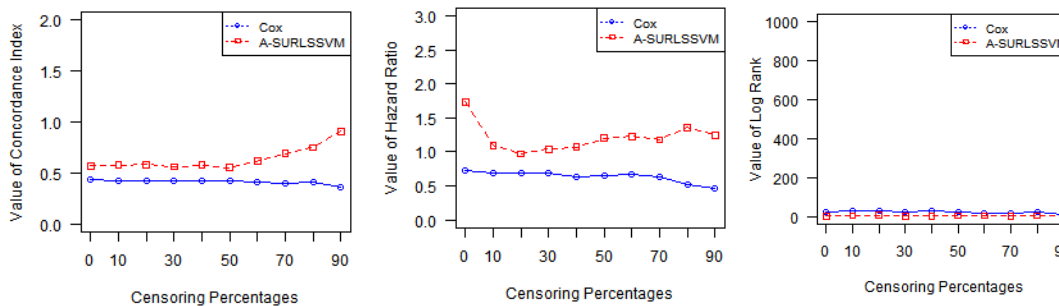
**Lampiran 41 b.** Performansi untuk Data Simulasi  $p=8$   $n=1000$  dengan Pola Linier



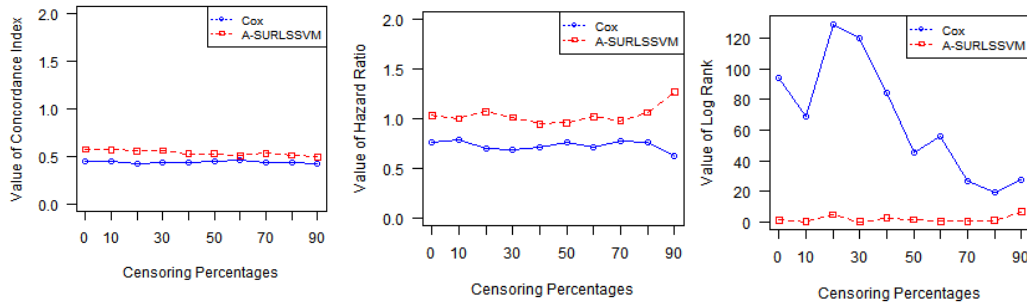
**Lampiran 41 c.** Performansi untuk Data Simulasi  $p=15$   $n=1000$  dengan Pola Linier



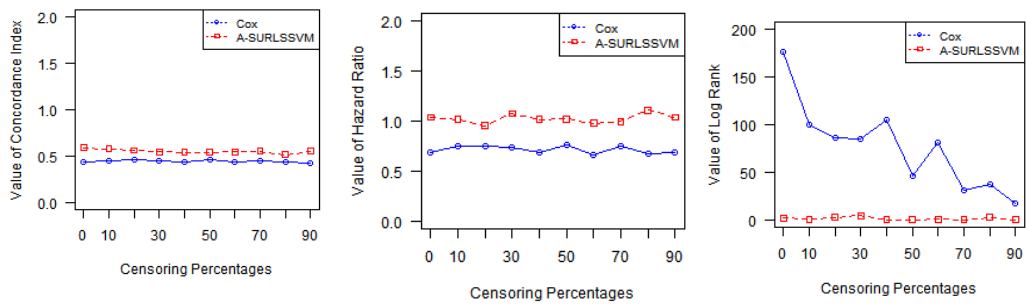
**Lampiran 41 d.** Performansi untuk Data Simulasi  $p=17$   $n=1000$  dengan Pola Linier



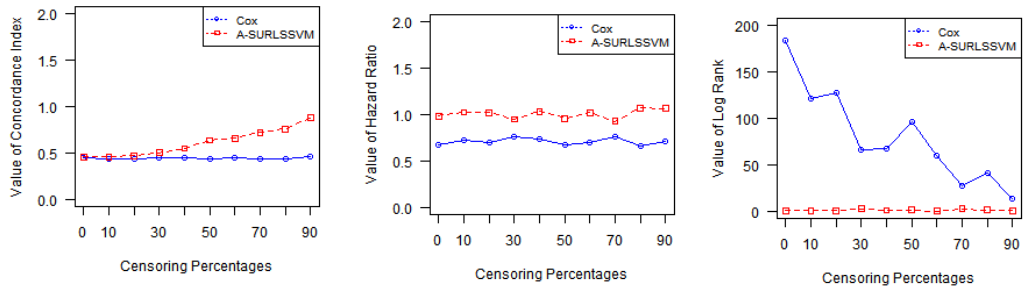
**Lampiran 42 a. Performansi untuk Data Simulasi  $p=6$   $n=5000$  dengan Pola Linier**



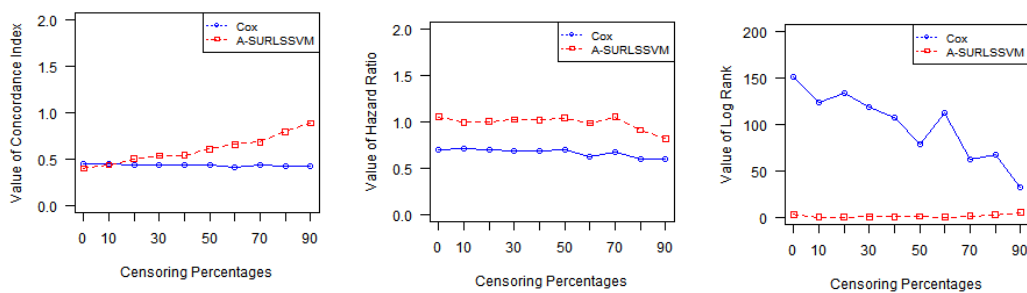
**Lampiran 42 b. Performansi untuk Data Simulasi  $p=8$   $n=5000$  dengan Pola Linier**



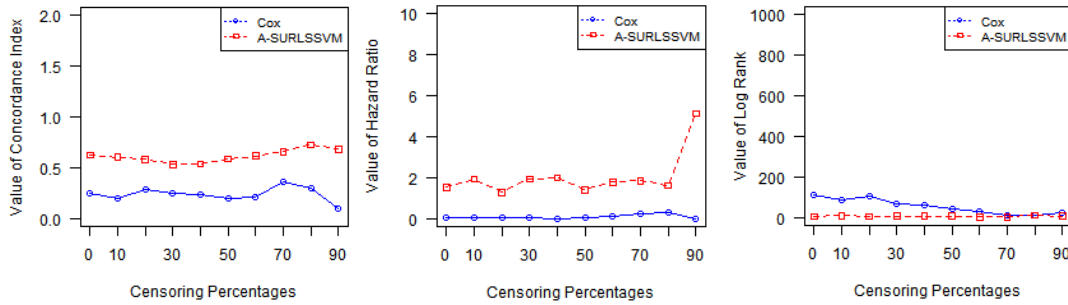
**Lampiran 42 c. Performansi untuk Data Simulasi  $p=15$   $n=5000$  dengan Pola Linier**



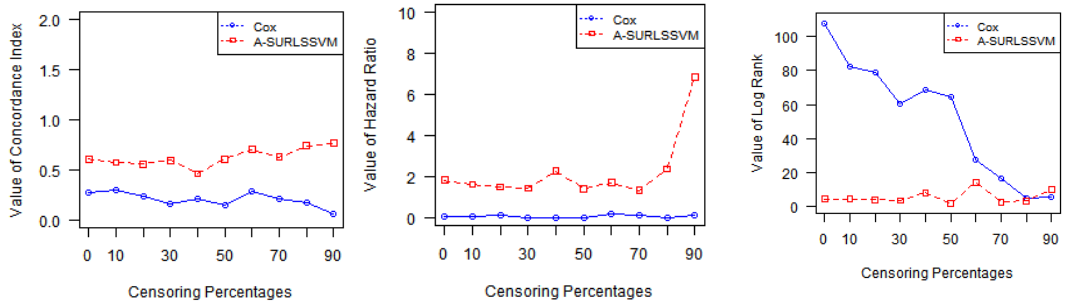
**Lampiran 42 d. Performansi untuk Data Simulasi  $p=17$   $n=5000$  dengan Pola Linier**



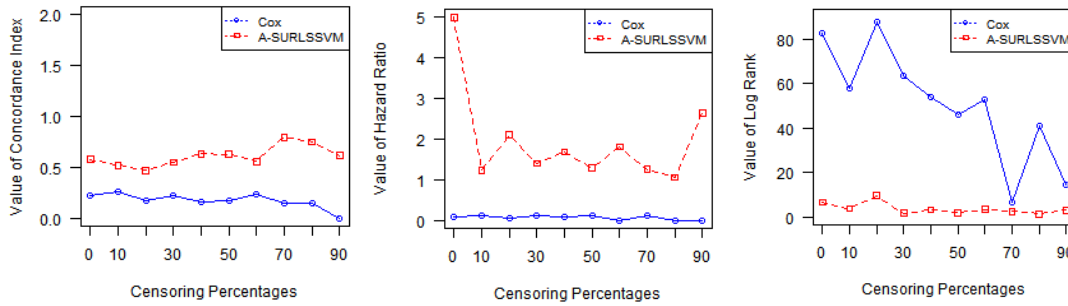
**Lampiran 43 a.** Performansi untuk Data Simulasi  $p=6$   $n=100$  dengan Pola Nonlinier



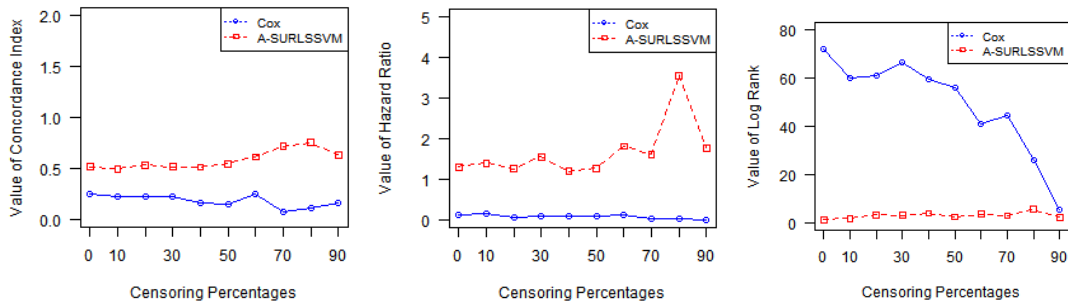
**Lampiran 43 b.** Performansi untuk Data Simulasi  $p=8$   $n=100$  dengan Pola Nonlinier



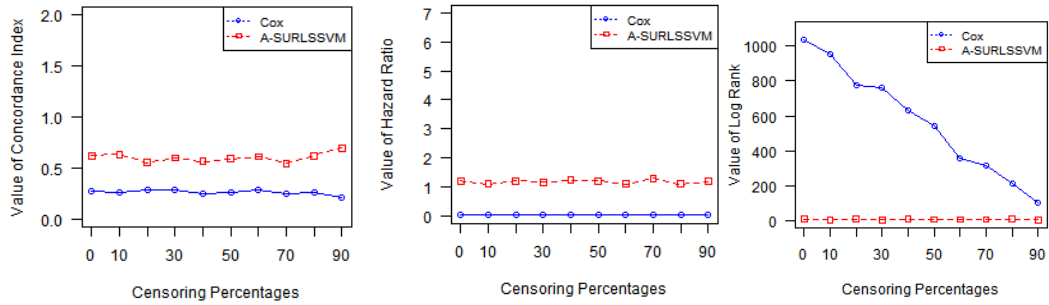
**Lampiran 43 c.** Performansi untuk Data Simulasi  $p=15$   $n=100$  dengan Pola Nonlinier



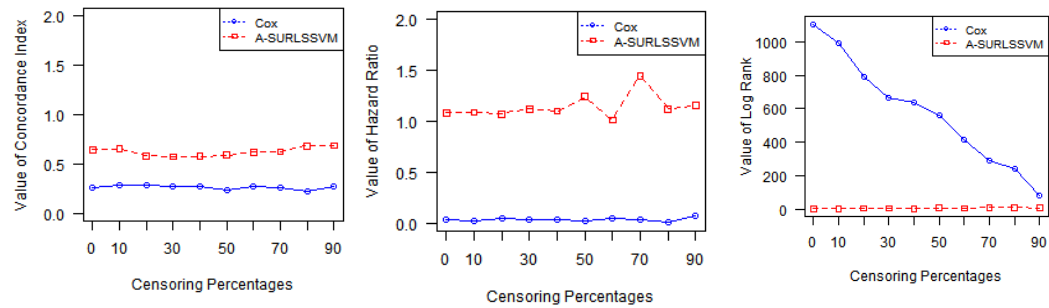
**Lampiran 44 d.** Performansi untuk Data Simulasi  $p=17$   $n=100$  dengan Pola Nonlinier



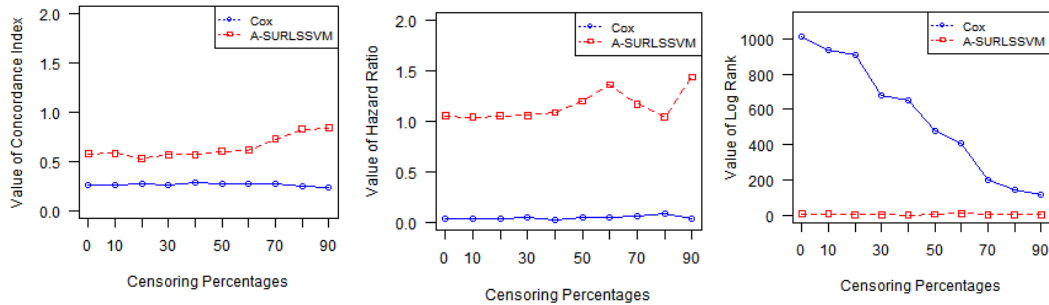
**Lampiran 44 a.** Performansi untuk Data Simulasi  $p=6$   $n=1000$  dengan Pola Nonlinier



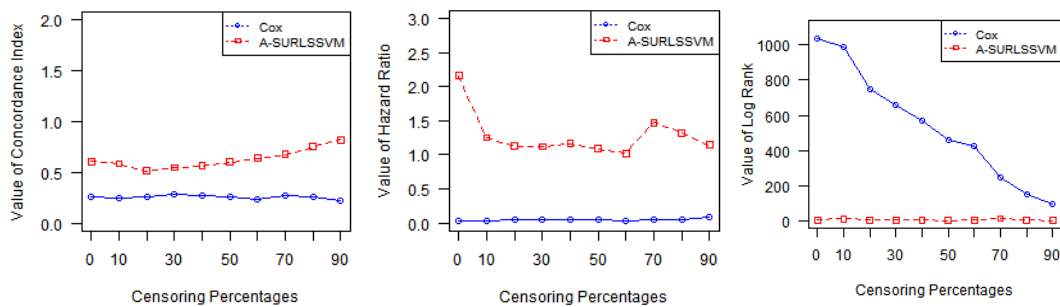
**Lampiran 44 b.** Performansi untuk Data Simulasi  $p=8$   $n=1000$  dengan Pola Nonlinier



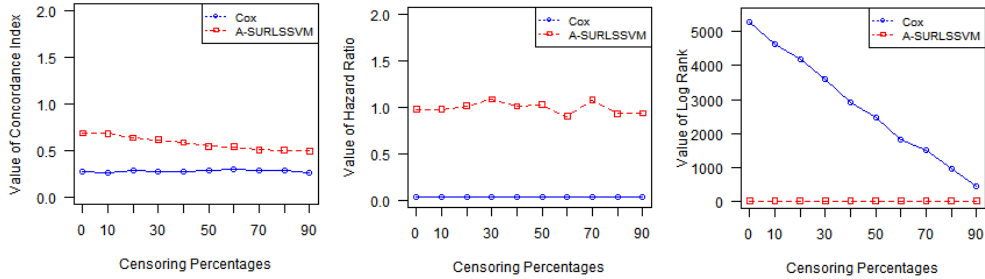
**Lampiran 44 c.** Performansi untuk Data Simulasi  $p=15$   $n=1000$  dengan Pola Nonlinier



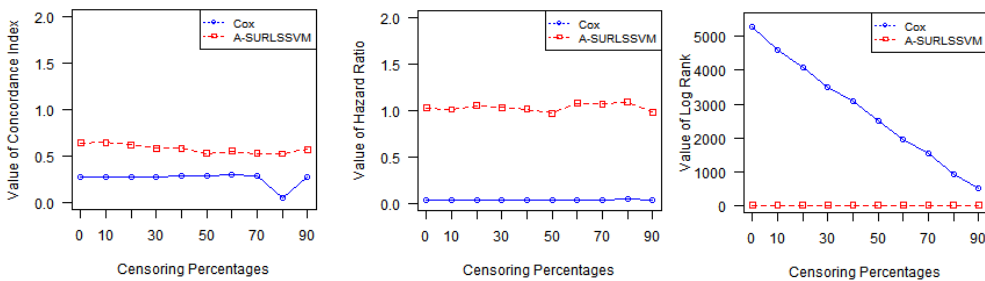
**Lampiran 44 d.** Performansi untuk Data Simulasi  $p=17$   $n=1000$  dengan Pola Nonlinier



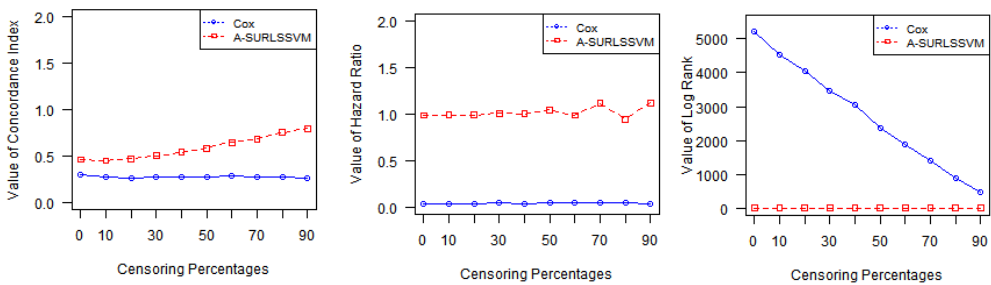
**Lampiran 45 a.** Performansi untuk Data Simulasi  $p=6$   $n=5000$  dengan Pola Nonlinier



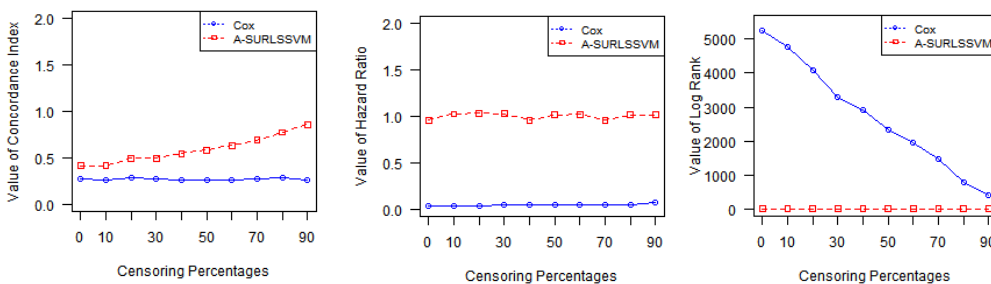
**Lampiran 45 b.** Performansi untuk Data Simulasi  $p=8$   $n=5000$  dengan Pola Nonlinier



**Lampiran 45 c.** Performansi untuk Data Simulasi  $p=15$   $n=5000$  dengan Pola Nonlinier



**Lampiran 45 d.** Performansi untuk Data Simulasi  $p=17$   $n=5000$  dengan Pola Nonlinier



**Lampiran 46.** Optimasi Parameter ASURLSSVM dan RBF kernel untuk Data Kanker Serviks

No	Hazard Ratio	Logrank	C-index	mu	C	kern_par
1	1.345717791	0.516537275	85.22613065	0.001	0.01	0.5
2	1.345717791	0.516537275	85.22613065	0.01	0.01	0.5
3	1.345717791	0.516537275	85.22613065	0.1	0.01	0.5
4	1.345717791	0.516537275	85.27638191	1	0.01	0.5
5	1.345717791	0.516537275	85.27638191	10	0.01	0.5
6	1.345717791	0.516537275	84.8241206	0.001	0.01	0.25
7	1.345717791	0.516537275	84.92462312	0.01	0.01	0.25
8	1.345717791	0.516537275	84.92462312	0.1	0.01	0.25
9	1.345717791	0.516537275	84.92462312	1	0.01	0.25
10	1.345717791	0.516537275	84.92462312	10	0.01	0.25
11	0.96707916	0.006148673	74.07035176	0.001	0.01	1
12	0.96707916	0.006148673	74.07035176	0.01	0.01	1
13	0.96707916	0.006148673	74.07035176	0.1	0.01	1
14	0.96707916	0.006148673	74.0201005	1	0.01	1
15	0.96707916	0.006148673	74.0201005	10	0.01	1
16	0.468070558	2.81998466	74.87437186	0.001	0.01	2
17	0.468070558	2.81998466	74.87437186	0.01	0.01	2
18	0.465225858	2.867881657	74.87437186	0.1	0.01	2
19	0.476667863	2.680180292	74.67336683	1	0.01	2
20	0.476667863	2.680180292	74.52261307	10	0.01	2
21	0.48149518	3.118090968	70.30150754	0.001	0.01	4
22	0.48149518	3.118090968	70.30150754	0.01	0.01	4
23	0.468358292	3.367675984	70.35175879	0.1	0.01	4
24	0.47666968	3.203267841	70.25125628	1	0.01	4
25	0.470149916	3.331365879	70.35175879	10	0.01	4
26	1.338189531	0.496754426	85.07537688	0.001	0.1	0.5
27	1.338189531	0.496754426	85.07537688	0.01	0.1	0.5
28	1.338189531	0.496754426	85.07537688	0.1	0.1	0.5
29	1.338189531	0.496754426	85.02512563	1	0.1	0.5
30	1.338189531	0.496754426	85.07537688	10	0.1	0.5
31	1.338189531	0.496754426	84.8241206	0.001	0.1	0.25
32	1.338189531	0.496754426	84.8241206	0.01	0.1	0.25
33	1.338189531	0.496754426	84.8241206	0.1	0.1	0.25
34	1.338189531	0.496754426	84.87437186	1	0.1	0.25
35	1.338189531	0.496754426	84.87437186	10	0.1	0.25
36	1.428330489	0.731000102	83.06532663	0.001	0.1	1
37	1.428330489	0.731000102	83.06532663	0.01	0.1	1



No	Hazard Ratio	Logrank	C-index	mu	C	kern_par
38	1.428330489	0.731000102	83.16582915	0.1	0.1	1
39	1.428330489	0.731000102	83.26633166	1	0.1	1
40	1.428330489	0.731000102	83.26633166	10	0.1	1
41	0.783943805	0.326657609	74.67336683	0.001	0.1	2
42	0.783943805	0.326657609	74.72361809	0.01	0.1	2
43	0.783943805	0.326657609	74.67336683	0.1	0.1	2
44	0.779857422	0.340694811	74.52261307	1	0.1	2
45	0.779857422	0.340694811	74.32160804	10	0.1	2
46	0.843859715	0.190432458	61.35678392	0.001	0.1	4
47	0.843859715	0.190432458	61.35678392	0.01	0.1	4
48	0.82898447	0.232576259	61.30653266	0.1	0.1	4
49	0.811619134	0.288618852	60.85427136	1	0.1	4
50	0.813866878	0.280930146	60.35175879	10	0.1	4
51	1.404978651	0.672652441	85.52763819	0.001	1	0.5
52	1.404978651	0.672652441	85.52763819	0.01	1	0.5
53	1.404978651	0.672652441	85.52763819	0.1	1	0.5
54	1.404978651	0.672652441	85.32663317	1	1	0.5
55	1.404978651	0.672652441	85.07537688	10	1	0.5
56	1.338189531	0.496754426	85.1758794	0.001	1	0.25
57	1.338189531	0.496754426	85.1758794	0.01	1	0.25
58	1.338189531	0.496754426	85.1758794	0.1	1	0.25
59	1.338189531	0.496754426	84.97487437	1	1	0.25
60	1.338189531	0.496754426	84.67336683	10	1	0.25
61	1.491718101	0.92545355	83.26633166	0.001	1	1
62	1.491718101	0.92545355	83.31658291	0.01	1	1
63	1.396980123	0.649587812	83.16582915	0.1	1	1
64	1.381168589	0.604855532	82.81407035	1	1	1
65	1.374034228	0.58505407	82.86432161	10	1	1
66	1.233228345	0.271943374	70.70351759	0.001	1	2
67	1.233228345	0.271943374	70.70351759	0.01	1	2
68	1.251057076	0.310533974	70.8040201	0.1	1	2
69	1.244664372	0.296297101	71.6080402	1	1	2
70	1.166422411	0.147026367	72.51256281	10	1	2
71	0.943786184	0.021666372	49.1959799	0.001	1	4
72	0.943786184	0.021666372	49.1959799	0.01	1	4
73	0.943786184	0.021666372	49.1959799	0.1	1	4
74	0.768086353	0.45050202	49.1959799	1	1	4
75	0.849832964	0.17417099	49.34673367	10	1	4
76	1.485521132	0.9134909	85.77889447	0.001	10	0.5

No	Hazard Ratio	Logrank	C-index	mu	C	kern_par
77	1.485521132	0.9134909	85.77889447	0.01	10	0.5
78	1.460685108	0.83700585	85.42713568	0.1	10	0.5
79	1.460685108	0.83700585	85.27638191	1	10	0.5
80	1.382267329	0.609347782	85.52763819	10	10	0.5
81	1.389758051	0.635186884	85.42713568	0.001	10	0.25
82	1.389758051	0.635186884	85.37688442	0.01	10	0.25
83	1.389758051	0.635186884	85.57788945	0.1	10	0.25
84	1.389758051	0.635186884	85.12562814	1	10	0.25
85	1.382267329	0.609347782	85.37688442	10	10	0.25
86	1.39124212	0.593671274	83.46733668	0.001	10	1
87	1.39124212	0.593671274	83.46733668	0.01	10	1
88	1.38005611	0.564238	83.31658291	0.1	10	1
89	1.632512572	1.402843567	82.86432161	1	10	1
90	1.380849741	0.604954438	82.31155779	10	10	1
91	1.228592132	0.231099021	70.8040201	0.001	10	2
92	1.228592132	0.231099021	70.75376884	0.01	10	2
93	1.228592132	0.231099021	70.65326633	0.1	10	2
94	1.577012013	1.273278155	70.50251256	1	10	2
95	1.689234479	1.797412057	70.70351759	10	10	2
96	0.911624474	0.055780188	49.94974874	0.001	10	4
97	0.911624474	0.055780188	49.94974874	0.01	10	4
98	0.911624474	0.055780188	49.84924623	0.1	10	4
99	1.107787887	0.070519974	48.69346734	1	10	4
100	1.001724095	3.68E-05	46.03015075	10	10	4
101	1.5165266	0.956302445	86.33165829	0.001	100	0.5
102	1.5165266	0.956302445	86.33165829	0.01	100	0.5
103	1.5165266	0.956302445	86.33165829	0.1	100	0.5
104	1.5165266	0.956302445	86.28140704	1	100	0.5
105	1.476286354	0.885631393	85.57788945	10	100	0.5
106	1.443139065	0.745654877	85.87939698	0.001	100	0.25
107	1.443139065	0.745654877	86.03015075	0.01	100	0.25
108	1.443139065	0.745654877	86.08040201	0.1	100	0.25
109	1.443139065	0.745654877	86.13065327	1	100	0.25
110	1.400080926	0.66483762	85.52763819	10	100	0.25
111	1.569937946	1.117357603	84.27135678	0.001	100	1
112	1.569937946	1.117357603	84.27135678	0.01	100	1
113	1.569937946	1.117357603	84.22110553	0.1	100	1
114	1.556610646	1.073898712	84.22110553	1	100	1
115	1.371568555	0.542202285	83.16582915	10	100	1

No	Hazard Ratio	Logrank	C-index	mu	C	kern_par
116	1.257944998	0.286977658	72.06030151	0.001	100	2
117	1.257944998	0.286977658	72.06030151	0.01	100	2
118	1.257944998	0.286977658	72.01005025	0.1	100	2
119	1.249354729	0.270258296	71.90954774	1	100	2
120	1.36236427	0.555590605	71.10552764	10	100	2
121	0.818605373	0.264780433	51.30653266	0.001	100	4
122	0.818605373	0.264780433	51.30653266	0.01	100	4
123	0.847896619	0.17906206	51.25628141	0.1	100	4
124	0.837432804	0.207154596	51.10552764	1	100	4
125	0.839357844	0.201711044	49.54773869	10	100	4

#### Lampiran 47. Syntax ASURLSSVM

```
asurlssvm = function(Y,X,delta,C,kern_par,mu) {  
  
  n = length(Y)  
  orderY = order(Y)  
  Ybaru = Y[orderY]  
  Xbaru = X[orderY,]  
  delta = delta[orderY]  
  
  source("rbf_kernel.R")  
  K = kernel_rbf(Xbaru, kern_par)  
  K = round(K,7)  
  
  source("makeDiffR.R")  
  D = makeDiff(Ybaru,delta)  
  D = as.matrix(D)  
  nD = dim(D)[1]  
  
  D2 = 0*D  
  D2[which(D==1)]=1  
  bc11 = -diag(n)-(mu*diag(delta)+C*t(D2)%*%D)%*%K  
  bc12 = (-C*(t(D2)%*%D+mu*diag(delta)))*%*%matrix(1,n,1)  
  bc21 = matrix(1,1,n)  
  bc22 = 0  
  
  A1 = data.frame(bc11,bc12)  
  A2 = data.frame(bc21,bc22)  
  A = rbind(A1, as.vector(as.matrix(A2)))  
  
  B = -C*(t(D2)%*%D)*%*%matrix(1,n,1)+mu*diag(delta)*%*%Ybaru)  
  B = rbind(B,0)  
  
  sol = solve(as.matrix(A),as.matrix(B))  
  alpha = sol[1:nD]  
  b = sol[length(sol)]  
  b = matrix(b,n,1)  
  u = t(alpha)%*%D)%*%K  
  alpha2 = t(D)%*%alpha  
  alpha2 = alpha2[orderY]  
  prognostik_indeks = u + t(b)  
  prognostik_indeks = t(prognostik_indeks)  
  list(alpha = alpha2, b = b, prognostik_indeks = prognostik_indeks)  
  
}
```

## Lampiran 48. Output A-SURLSSVM pada Data Pasien Kanker Serviks

Salpha					
[1]	-0.493601428	-0.470936704	0.427162946	-13.546231638	1.506969741
[6]	1.491913023	1.968619715	1.944643862	-0.472792981	1.735520537
[11]	1.959394541	0.421217804	-0.474423928	-0.834428682	1.717656028
[16]	-0.472835829	2.274392372	0.255032755	-0.472641731	-0.317097292
[21]	-0.472730423	-0.473636696	1.052858551	0.428919019	-0.472858094
[26]	-0.312297141	-0.311833265	1.959139089	-0.477583676	-0.472799619
[31]	-0.474641613	0.427165816	-0.472498619	-0.469414591	-0.508550693
[36]	-0.392464021	-0.311988849	-0.181166133	-0.042867178	0.637226299
[41]	1.491912966	1.942315315	1.853121288	-0.472857972	-0.297354767
[46]	1.266782227	-0.498535214	-0.312029416	-0.312758257	0.266060371
[51]	-0.310842657	-0.312082240	1.960233212	-0.472854470	-0.472877732
[56]	-0.472854842	1.281179012	-0.472850285	-0.467738254	-0.319066468
[61]	-0.472162977	-0.472850182	-0.472838837	0.427257654	1.268123204
[66]	-0.472854922	-0.474141965	-0.472862928	-0.493554313	0.427681012
[71]	-0.472643317	-0.311949586	0.426829366	-5.944872233	-0.474247668
[76]	-0.472270641	-0.433424522	2.160546160	-0.472858094	-0.508556645
[81]	25.144597554	-0.316326302	-0.311881066	0.431436479	-0.472850275
[86]	-0.472863624	0.427316116	-0.576700187	0.427204739	-0.472716463
[91]	-0.472780605	0.254839562	0.424199391	1.959395888	-0.472851652
[96]	-0.464371709	-0.464276552	-0.464493688	-0.469657043	-0.473597981
[101]	-2.767787210	0.427149907	0.427894733	-0.530035663	-0.493282580
[106]	0.254933754	-0.492012814	-0.446843904	1.717455030	-0.312030411
[111]	1.267788795	-0.315134355	-0.311574328	2.251873099	0.578065204
[116]	-0.472861528	-0.472858094	-0.472858296	-0.312012796	-0.312661694
[121]	0.249724983	-0.511582196	-0.493600913	-0.472857439	-0.472765189
[126]	-0.311993275	-0.312012796	0.427263973	0.427768825	1.159145352
[131]	-0.472823437	-0.473266094	-0.311471621	-0.311867135	-0.312012796
[136]	0.427144249	-0.472820084	-0.471605646	-0.472914721	-0.312012415
[141]	-0.312035912	0.427146051	-2.531628918	2.251872725	-0.464692667
[146]	-0.474294039	-0.316947727	-0.049641201	2.251872624	-0.472835675
[151]	-0.303724504	-0.311989209	-0.303525426	1.979053644	-0.555121746
[156]	0.050585667	1.959345299	0.833629826	-32.854217948	-0.472857015
[161]	-0.472604087	-0.311909196	-0.311996802	0.254681437	1.072295387
[166]	2.251872808	-0.047857769	0.427162946	-0.574498933	-0.472858094
[171]	-0.312729782	-0.464125112	0.259130101	1.959203020	-0.050605717
[176]	-0.472857426	-0.312012796	0.255943014	0.427923661	-0.511572074
[181]	-0.005527931	-0.047408741	-0.472878658	-0.467200968	0.427052459
[186]	0.427129290	1.720017430	-0.472744601	-0.424357302	1.959296451
[191]	-0.511581083	-0.471341286	-0.312012734	0.427153991	-0.472179405
[196]	-0.311805875	2.251285823	-0.472784436	-0.311431634	-0.472469421
[201]	-0.314065822	1.305607700	-0.050586891	-0.050009980	0.102966040
[206]	-0.472819029	-0.312011600	-0.041440169	0.842066517	1.717881375
[211]	-0.473189952	1.717978368	-0.493602299	-0.476092029	-0.055239637
[216]	1.968683913	1.579805412	0.255058530	-0.312011971	-0.472629023
[221]	-0.312011673	-0.472858094	-0.472771532	-0.297354580	-0.312009865
[226]	-0.472869910	-0.050610805	0.427162946	0.576255001	-0.476486476
[231]	-0.311837883	-0.312012761	0.019354171	0.833959834	-0.474405503
[236]	-0.312012796	0.331461692	-0.472834842	-0.312010756	-0.038358457
[241]	0.427283984	-0.469548469	1.717464833	-0.472658089	-0.050583389
[246]	-0.392531361	-0.304396359	0.427162946	2.251784531	-0.470133023
[251]	-2.941261317	1.959394541	1.167743460	-0.472989430	-0.312156755
[256]	0.260464690	0.427162946	1.719643162	-0.472858094	0.015657427
[261]	-0.050605186	-0.472805060	-0.472858221	-0.470719124	-0.470972868
[266]	-0.472856866	-0.578766071	-0.311993650	-0.472320338	-0.311996245
[271]	0.427165525	-0.311896903	0.631092198	-0.493600913	-0.311990569
[276]	-0.472866902	-0.472787000	0.427694364	-0.311558601	-0.049834865
[281]	-0.472812001	-0.469102430	-0.311647601	-0.493968354	-0.312015664
[286]	-0.472806348	-0.472898070	-0.310273965	-0.312012796	1.958866915
[291]	-0.469239502	-0.493447331	-0.472858094	2.252672253	-0.017829201
[296]	-0.472857953	-0.472654262	-0.050674248	69.523527903	-0.472835912

**Lampiran 48.** Output A-SURLSSVM pada Data Pasien Kanker Serviks

```

[301] -0.472728580 -0.470437749 -0.049475292 -0.472857291 -0.471807044
[306] 0.104070945 -0.472952902 -0.582454121 1.338145529 0.429548609
[311] -0.472547510 -0.472700794 0.257839455 1.959416298 -0.027365811
[316] -0.312015912 -0.472863688 -0.312030215 -0.156781082 -0.050600669
[321] -0.491441793 -0.472858229 -0.050630741 -0.050639315 -0.472764366
[326] 0.991117874 0.254998719 -0.311999377 -0.543513796 -0.472858094
[331] -0.472857880 -0.322896849 7.929120563 -0.472097623 -0.472878750
[336] -0.469915694 -0.312376426 -0.472857845 -0.472599771 -0.472824474
[341] -0.181870812 -0.472901151 -0.472863118 -0.322735749 -0.472858094
[346] -0.184120246 -0.182230441 -0.472942625 0.427700320 -0.493475340
[351] 0.427278094 -14.530756031 0.427112559 -0.311246977 -0.472857336
[356] -0.050600044 -0.311988786 -0.472857330 -0.472886492 -0.473816940
[361] -0.472300421 -0.308266224 -0.472443314 -0.472858094 -0.312012796
[366] -0.313668381 -0.312012796 -0.472851532 -0.472419716 -23.993559458
[371] -0.493600820 -0.311833993 0.000000000 -0.471696568 -0.312006951
[376] -0.470788974 -0.312041853 -0.312012603 -0.475575317 -0.312004679
[381] -0.464182038 -0.474385185 -0.312091366 -0.472840120 0.000000000
[386] -0.473241613 -0.473094547 -0.472858094 -0.473093474 -0.022798658
[391] -0.493484731 -0.476466783 -0.156781082 -0.311584981 -0.472791932
[396] -0.312112822 -0.311997243 -0.472864382 -0.576441318 -0.493453574
[401] -0.472695937 -0.472849286 -0.308361948 -0.472858094 0.000000000
[406] -0.472694580 -0.471511548 1.959396851 -0.494166706 0.000000000
[411] 0.000000000 0.000000000

```

\$b

[,1]

[1,] 8.830366

\$prognostik\_indeks

[,1]

```

42 8.831710
116 8.831219
337 8.830059
373 8.830363
374 8.830366
375 8.830379
393 9.990967
390 8.248638
141 9.415618
377 8.253339
398 9.408524
356 8.238674
100 9.997948
362 8.255867
379 8.227280
92 8.812210
404 8.824836
317 8.807567
52 8.849717
214 10.410153
109 8.275244
159 8.286843
367 8.300331
353 16.759474
50 8.318091
221 8.318794
265 8.318809
396 8.331708
45 8.338185
91 8.336765
164 8.336200
343 8.336549
82 8.336811
88 8.336613
101 8.343194
195 8.337035
260 8.336447

```

**Lampiran 49.** Prognostik indeks pada Data Pasien Kanker Serviks

No	prognostik	T	No	prognostik	T	No	prognostik	T
1	11.08225	712	41	8.786905	159	81	8.357333	51
2	11.08607	625	42	8.83171	3	82	8.336811	25
3	11.08224	505	43	8.51836	138	83	8.519756	93
4	11.17918	361	44	8.833087	147	84	8.780723	160
5	-4.71675	254	45	8.338185	24	85	8.781526	156
6	10.53955	277	46	9.256729	212	86	8.523514	129
7	8.357489	66	47	9.257513	206	87	-24.0239	273
8	8.930475	170	48	8.406001	221	88	8.336613	25
9	8.518354	105	49	8.518393	101	89	8.532421	111
10	8.360928	50	50	8.318091	23	90	9.161299	193
11	8.355233	64	51	8.351869	65	91	8.336765	24
12	10.90158	325	52	8.849717	17	92	8.81221	12
13	8.513897	116	53	8.778627	155	93	10.54742	271
14	8.357467	74	54	11.08196	429	94	9.257517	206
15	8.510026	131	55	8.36576	47	95	8.357513	52
16	8.765874	155	56	8.353899	68	96	8.357501	34
17	8.356668	80	57	10.82433	278	97	9.090609	172
18	10.77511	287	58	10.55001	276	98	9.259918	205
19	8.518337	107	59	8.359421	64	99	8.359402	72
20	8.518707	97	60	11.10476	347	100	9.997948	7
21	8.355506	70	61	10.09712	247	101	8.343194	25
22	9.257486	198	62	8.357502	87	102	8.781666	152
23	8.357919	57	63	8.360712	39	103	8.437901	88
24	9.258135	205	64	8.356069	34	104	9.662297	240
25	8.357693	87	65	8.336397	26	105	8.366312	53
26	8.358944	79	66	8.518877	122	106	8.359254	65
27	8.357603	65	67	9.257531	208	107	8.517801	91
28	8.357507	27	68	9.467312	230	108	8.358836	31
29	8.357547	59	69	9.257622	208	109	8.275244	18
30	8.357483	65	70	8.513268	96	110	8.357679	77
31	10.99091	445	71	8.524234	135	111	9.257588	206
32	9.25806	214	72	8.321809	64	112	8.518357	129
33	8.779764	153	73	8.518176	108	113	8.357423	46
34	10.09816	251	74	8.518203	95	114	8.34742	45
35	8.518266	96	75	8.356957	40	115	10.78923	304
36	8.365744	57	76	8.78089	160	116	8.831219	3
37	10.7903	288	77	8.520605	126	117	8.357489	47
38	-15.1632	327	78	8.357509	32	118	10.79039	321
39	9.182933	189	79	8.389277	53	119	8.518327	115
40	8.338661	42	80	8.357677	86	120	8.354334	48

**Lampiran 49.** Prognostik indeks pada Data Pasien Kanker Serviks (Lanjutan)

No	prognostik	T	No	prognostik	T	No	prognostik	T
121	9.254681	200	161	8.649184	144	201	10.78966	295
122	8.357514	67	162	10.78892	332	202	8.64821	146
123	9.085365	189	163	9.461219	235	203	8.357341	66
124	10.11149	246	164	8.3362	24	204	9.898155	243
125	8.357287	67	165	8.523014	66	205	-5.70039	187
126	8.36579	67	166	8.517554	105	206	8.792007	149
127	8.511306	109	167	10.7904	286	207	9.257419	205
128	8.355909	85	168	8.35667	53	208	10.799	291
129	8.517599	100	169	8.779674	149	209	8.359286	69
130	8.356331	66	170	8.788906	160	210	8.357516	69
131	2.88506	241	171	8.357542	38	211	9.081396	182
132	8.518366	116	172	8.359298	48	212	10.54614	267
133	8.783062	167	173	8.880902	165	213	10.55038	274
134	8.779726	149	174	8.518284	125	214	10.41015	17
135	10.33734	254	175	8.35749	83	215	8.35758	66
136	8.357338	75	176	10.77328	331	216	8.437854	88
137	8.517206	141	177	9.258534	200	217	8.357485	44
138	9.257204	205	178	8.357803	44	218	8.360233	64
139	8.357834	52	179	8.648544	144	219	8.394156	62
140	9.08631	172	180	8.517784	139	220	8.357509	64
141	9.415618	5	181	8.355198	53	221	8.318794	23
142	8.518336	104	182	8.779738	165	222	5.889004	222
143	8.51815	112	183	8.357476	64	223	8.3575	35
144	8.35756	51	184	8.358065	68	224	8.360074	72
145	10.79086	290	185	8.522193	102	225	8.518218	106
146	9.672432	238	186	8.35751	59	226	9.085046	184
147	9.25737	228	187	8.357502	69	227	9.258066	203
148	8.522099	125	188	11.08304	428	228	8.357413	65
149	8.359928	74	189	8.932499	170	229	8.518069	100
150	10.78892	284	190	9.257639	205	230	8.518291	109
151	8.520071	94	191	8.357885	72	231	8.357508	29
152	8.359024	38	192	8.357479	40	232	8.518352	103
153	10.32168	262	193	8.357061	47	233	8.358089	77
154	9.085373	184	194	8.359488	46	234	9.257529	220
155	10.54547	263	195	8.337035	25	235	8.349916	66
156	8.365673	33	196	8.518928	113	236	10.80291	319
157	10.78974	281	197	8.355567	62	237	8.356756	72
158	9.254565	210	198	8.357288	27	238	8.359831	69
159	8.286843	20	199	8.35753	67	239	8.519523	105
160	8.776898	153	200	10.56587	262	240	8.518902	136



**Lampiran 49.** Prognostik indeks pada Data Pasien Kanker Serviks (Lanjutan)

No	prognostik	T	No	prognostik	T	No	prognostik	T
241	8.51835	135	281	8.336912	25	321	8.518328	97
242	8.518437	102	282	9.260631	206	322	9.256976	194
243	8.358632	61	283	9.259318	211	323	8.353237	47
244	8.365361	67	284	8.518243	122	324	9.258418	214
245	8.517901	94	285	8.519233	93	325	8.518513	108
246	8.517854	92	286	9.257544	192	326	8.484086	117
247	9.821484	138	287	8.771266	157	327	8.366081	66
248	10.09849	253	288	8.518491	109	328	8.363292	65
249	10.68349	341	289	9.261822	190	329	8.35461	43
250	10.32246	254	290	9.096309	187	330	9.256419	218
251	10.78951	305	291	8.514589	112	331	9.256871	213
252	9.883488	244	292	9.251402	221	332	8.357544	38
253	8.357531	65	293	8.506322	104	333	8.357417	65
254	9.257495	195	294	9.080136	180	334	9.091026	171
255	8.515426	134	295	8.518055	96	335	8.353775	44
256	8.352782	26	296	8.35395	76	336	8.355922	48
257	8.357549	74	297	8.357264	73	337	8.830059	3
258	9.665292	242	298	10.06793	142	338	8.36103	53
259	8.646178	142	299	8.338924	25	339	8.396959	87
260	8.336447	25	300	8.777183	160	340	8.352283	50
261	8.355781	47	301	8.357553	68	341	8.518368	109
262	7.985337	239	302	8.673577	94	342	8.525969	135
263	8.349231	29	303	8.205109	93	343	8.336549	24
264	8.359407	47	304	8.357511	70	344	8.356945	67
265	8.318809	23	305	8.513859	122	345	8.365873	51
266	8.525914	98	306	8.313297	53	346	8.518532	114
267	8.518176	108	307	8.330115	82	347	8.357575	85
268	8.205109	93	308	6.059708	165	348	8.380163	85
269	8.357498	68	309	8.66512	93	349	8.357946	36
270	8.779227	147	310	8.532988	117	350	8.517922	141
271	8.518009	113	311	8.3576	58	351	8.35811	53
272	8.517934	117	312	8.357508	75	352	8.361263	78
273	9.257545	208	313	8.356237	78	353	16.75947	22
274	33.97497	87	314	8.35505	51	354	8.518797	104
275	9.257561	194	315	8.351938	64	355	10.16764	86
276	8.357508	27	316	9.086309	172	356	8.238674	6
277	8.779744	149	317	8.807567	15	357	8.518226	115
278	10.54782	262	318	8.35247	72	358	8.360714	73
279	9.257031	210	319	9.26327	217	359	8.337212	25
280	8.520133	93	320	8.354991	28	360	8.50763	136

**Lampiran 49. Prognostik indeks pada Data Pasien Kanker Serviks (Lanjutan)**

No	prognostik	T	No	prognostik	T
361	8.526703	94	387	8.503241	115
362	8.255867	10	388	11.08224	436
363	8.510654	107	389	8.519214	96
364	8.357195	44	390	8.248638	4
365	8.357499	71	391	9.085957	177
366	8.35714	72	392	8.357451	29
367	8.300331	20	393	9.990967	3
368	8.356681	33	394	8.401834	86
369	8.357423	26	395	8.756231	150
370	8.357707	67	396	8.331708	23
371	8.357465	61	397	8.357276	75
372	8.357507	38	398	9.408524	5
373	8.830363	3	399	8.356968	51
374	8.830366	3	400	6.298606	250
375	8.830379	3	401	8.33685	25
376	78.35389	25	402	8.357223	70
377	8.253339	5	403	8.336765	25
378	8.356916	59	404	8.824836	14
379	8.22728	10	405	8.514739	88
380	8.518539	95	406	8.802226	154
381	8.518313	111	407	8.357685	67
382	8.357538	71	408	8.356124	62
383	8.357282	85	409	8.357121	46
384	8.514533	117	410	8.355382	68
385	8.522793	94	411	8.522004	114
386	8.357502	69	412	8.351371	72

*(halaman ini sengaja dikosongkan)*

## BIOGRAFI PENULIS



Penulis dengan nama lengkap Chusnul Khotimah, lahir di Tulungagung 24 tahun silam. Anak kedua dari Suwono dan Surtini serta adik dari Deni Kristanto. Selama hidupnya, penulis telah menempuh pendidikan di SDN Kepuh II tahun 1999-2005, SMPN 1 Tulungagung tahun 2005-2008, SMAN 1 Kedungwaru tahun 2008-2011. Setelah lulus SMA, penulis melanjutkan study S1 dan S2 di Departemen Statistika ITS.

Perjuangan adalah motivasi penulis selama menempuh pendidikan khususnya pada masa kuliah. Selama kuliah S1, penulis aktif di berbagai organisasi. Penulis bergabung di dua organisasi yaitu menjadi Staff Sosial Masyarakat HIMASTA-ITS 12/13 serta bergabung di BSO IECC BEM ITS 12/13. Di tahun ketiganya, penulis diamanahi sebagai Kadep SOSMAS HIMASTA-ITS 13/14. Selain itu penulis juga aktif di berbagai kepanitiaan, salah satunya adalah menjadi koordinator tim soal STATION 2013. Penulis juga aktif di dunia kepanitiaan, dari LKMM pra TD hingga menjadi Fasilitator dan Pemandu LKMM TM. Ada beberapa prestasi yang Alhamdulillah telah dicapai oleh penulis selama masa kuliah. Diantaranya adalah juara 2 Calculus Cup UNJ 2012, Juara 2 *National Statistics Challenge* (NSC) UB 2014, 1 PKM GT dan 5 PKM-P mendapatkan dana hibah dari Dikti. Keinginan menjadi dosen, mendorong penulis untuk melanjutkan studi S2. Dengan beasiswa LPDP, alhamdulillah penulis dapat melanjutkan S2. Selama S2 penulis aktif menjadi asisten dosen pada mata kuliah biostatistik dan analisis multivariat. Selain itu, sebagai awardee LPDP penulis juga aktif ikut serta dalam kegiatan-kegiatan yang ada. Untuk informasi maupun saran dari tesis ini, pembaca dapat menghubungi penulis di [chusnul.its2011@gmail.com](mailto:chusnul.its2011@gmail.com) atau facebook penulis yaitu chusnul khotimah. Penulis juga dapat dihubungi di 085736958148.

*(halaman ini sengaja dikosongkan)*