



TESIS - SS142501

ANALISIS DATA SURVIVAL DENGAN SURVIVAL LEAST SQUARE-SVM

(Studi Simulasi dan Studi Kasus Pasien Kanker Serviks)

HALWA ANNISA KHOIRI
NRP. 06211550010201

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

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



THESIS - SS142501

SURVIVAL ANALYSIS USING SURVIVAL LEAST SQUARE-SVM

(Case Study: Simulated Study and Cervical Cancer Patients)

HALWA ANNISA KHOIRI
NRP. 06211550010201

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

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

ANALISIS DATA SURVIVAL DENGAN SURVIVAL LEAST SQUARE-SVM
(Studi Simulasi dan Studi Kasus Pasien Kanker Serviks)

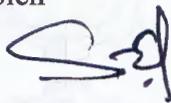
Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M.Si)

di
Institut Teknologi Sepuluh Nopember
Oleh:

HALWA ANNISA KHOIRI
NRP: 06211550010201

Tanggal Ujian : 11 Januari 2018
Periode Wisuda : Maret 2018

Disetujui oleh



1. Santi Wulan Purnami, M.Si, Ph.D (Pembimbing I)
NIP. 19720923 199803 2 001



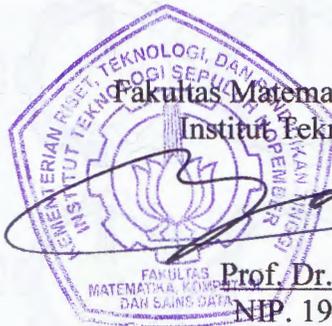
2. Dr. rer. pol. Dedy Dwi Prastyo, M.Si (Pembimbing II)
NIP. 19831204 200812 1 002



3. Dr. Bambang Widjanarko Otok, M.Si (Penguji I)
NIP. 19681124 199412 1 001

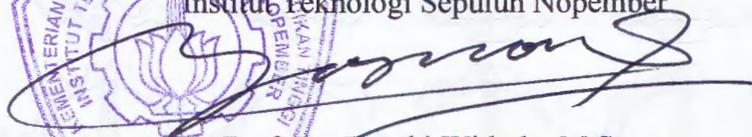


4. Dr. Suhartono, M.Sc (Penguji II)
NIP. 19710929 199512 1 001



Dekan

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



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

ANALISIS DATA SURVIVAL DENGAN SURVIVAL LEAST SQUARE-SVM (Studi Simulasi dan Studi Kasus Pasien Kanker Serviks)

Nama Mahasiswa : Halwa Annisa Khoiri
NRP : 06211550010201
Dosen Pembimbing : Santi Wulan Purnami, M.Si, Ph.D
Dr.rer.pol. Dedy Dwi Prastyo, M.Si

ABSTRAK

Dalam analisis data survival terdapat tiga pendekatan yang dapat digunakan, yaitu pendekatan parametrik, semi-parametrik, dan non-parametrik. Pendekatan parametrik memiliki kelemahan karena harus diketahui distribusi dari data *survival*. Pendekatan semi-parametrik salah satunya yang paling banyak digunakan adalah Cox Proportional Hazard Model. Model ini memiliki kelemahan karena harus memenuhi asumsi PH. Sehingga, dikembangkan metode dengan pendekatan non-parametrik yaitu Survival Support Vector Machine. Namun Survival-SVM memiliki kelemahan karena bentuk fungsinya adalah kuadratik sehingga membutuhkan komputasi yang lebih rumit untuk data berdimensi tinggi. Selanjutnya muncul penelitian baru yaitu Survival Least Squares-SVM (SURLS-SVM). Metode ini tidak terikat dengan asumsi distribusi sehingga lebih fleksibel dan kendalanya dapat diselesaikan secara linear sehingga membutuhkan komputasi yang lebih sederhana. SURLS-SVM memiliki variabel respon yang bersifat kontinu, dan langkah-langkah analisis dengan SURLS-SVM adalah sebagai berikut, pertama objek diurutkan berdasarkan *survival time* dan status tersensor, selanjutnya menghitung nilai *prognostic index* yang merupakan pengganti fungsi hazard pada Cox PHM berdasarkan data yang sudah diurutkan, dan terakhir membandingkan urutan *survival time* dengan prediksi *prognostic index*. Pada penelitian ini SURLS-SVM diaplikasikan untuk menganalisis kasus kanker serviks di RS dr. Soetomo Surabaya. Sebagai perbandingan juga dilakukan studi simulasi dengan ukuran data *survival* yang relatif besar. Berdasarkan hasil simulasi maupun data real, model SURLS-SVM lebih unggul dari Cox PHM terutama jika dilihat berdasarkan nilai c-index. Untuk meningkatkan performansi c-index, dilakukan *feature selection* dengan menggunakan *backward elimination* dan pada data simulasi dapat menaikkan c-index sampai 10%. Pada data real dapat digunakan untuk mengetahui variabel yang signifikan mempengaruhi c-index, variabel tersebut adalah usia, status komplikasi, usia saat menikah pertama, paritas, stadium, tingkat pendidikan, siklus menstruasi, jenis pengobatan, dan lama menstruasi.

Kata kunci: Analisis *Survival*, Cox PHM, Kanker Serviks, Survival Least Squares SVM.

(Halaman ini sengaja dikosongkan)

**SURVIVAL DATA ANALYSIS USING
SURVIVAL LEAST SQUARE-SVM**
(Case Study: Simulated Study and Cervical Cancer Patients)

Name : Halwa Annisa Khoiri
NRP : 06211550010201
Supervisor : Santi Wulan Purnami, M.Si, Ph.D
Dr.rer.pol. Dedy Dwi Prastyo, M.Si

ABSTRACT

The survival analysis has three approaches, i.e. parametric, semi-parametric, and non-parametric. Parametric approach has weakness that it should be known priori the distribution of survival time. Meanwhile, semi-parametric approach does not require priori known distribution of survival time, but this approach must fulfill proportional hazard assumption. The Cox Proportional Hazard Model (Cox PHM) is the popular semi-parametric model used to analyse survival data. One of the non-parametric approach is Survival Support Vector Machine (SUR-SVM). This model has disadvantage because the constrain function has quadratic form so that it requires more complicated computation for high-dimensional data. To overcome this drawback, Survival Least Square Support Vector Machine (SURLS-SVM) is developed. This model does not require distribution and PH assumption, and the constrain function is equality form that can be solved by linear programming. The SURLS-SVM has continuous response, and the algorithm as follows. First, objects are ranked based on survival time and censored status, then prognostic index instead of hazard function in Cox PHM is calculated using sorted data, and the last the survival time is compared with prediction of prognostic index. This research apply on cervical cancer data in dr. Soetomo Hospital Surabaya. The simulation study also uses this model for bigger sample size. Based on simulation result and real data result, SURLS-SVM outperforms than Cox PHM (from c-index value). The increasing of c-index value can be obtained by feature selection, i.e. backward elimination. In simulation study, the c-index increase 10%. In real dataset, this method can be used to know significant variable, i.e. age, complication status, age of first married, parity, stadium, level of education, menstruation cycle length, type of treatment, and length of menstruation.

Keywords: Survival Analysis, Cox PHM, Cervical Cancer, Survival Least Squares SVM.

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT, karena atas segala rahmat dan ridho-Nya sehingga tesis yang diberi judul “Analisis Data Survival dengan Survival Least Square-SVM (Studi Simulasi dan Studi Kasus Pasien Kanker Serviks)” ini bisa terselesaikan dengan baik. Tesis ini merupakan salah satu syarat untuk menyelesaikan pendidikan di Program Magister S2 Statistika ITS. Ada banyak pihak yang telah membantu dalam penulisan tesis ini, sehingga penulis ingin menyampaikan ucapan terima kasih kepada

1. Ibu Santi Wulan Purnama, M.Si, Ph.D dan Bapak Dr. rer. pol. Dedy Dwi Prastyo, M.Si selaku dosen pembimbing, yang telah bersedia meluangkan waktu untuk memberikan bimbingan, saran, dan ilmu yang sangat bermanfaat dalam penulisan tesis ini.
2. Bapak Dr. Bambang Wijanarko Otok, M.Si dan Bapak Dr. Suhartono, M.Sc, selaku dosen penguji yang telah memberikan banyak kritik dan saran sehingga tesis ini menjadi lebih baik.
3. Bapak Dr. Suhartono, M.Sc. selaku Kepala Departemen Statistika FMKSD ITS dan Bapak Dr. rer. pol. Heri Kuswanto, M.Si. selaku Kaprodi Pascasarjana Statistika FMKSD ITS.
4. Bapak /Ibu dosen pengajar di Departemen Statistika ITS, terima kasih atas semua ilmu berharga yang telah diberikan.
5. Bapak/Ibu staf dan karyawan di Departemen Statistika ITS, terima kasih atas segala bantuan selama masa perkuliahan penulis.
6. LPDP yang telah memberikan dukungan terutama dalam hal pembiayaan perkuliahan penulis.
7. Kedua orang tua yang sangat penulis hormati dan sayangi. Terima kasih karena telah menjadi alasan penulis untuk tidak menyerah dalam menyelesaikan tesis ini. Terima kasih atas semua doa dan perhatian yang selalu diberikan kepada penulis. Terima kasih juga untuk adik (Faricha Khairunnafi) yang selalu memberikan semangat kepada penulis dalam menyelesaikan tesis ini.

8. Calon suami (Johan Nur W.) yang telah memberikan semangat dan dukungan. Terima kasih sudah meluangkan waktu menjadi tempat berbagi penulis selama menyelesaikan tesis ini.
9. Teman-teman seperjuangan menyelesaikan tesis ini (Chusnul, Ayya, mbak Amel) terima kasih sudah mendengarkan segala keluh kesah dan bersedia menjadi tempat berbagi penulis selama menyelesaikan tesis ini.
10. Teman-teman Statistika 2015 Genap (Fani, Nisa, Ribka, Titis, Sem, alm.Mail, Lutfi, Mas Rama, Jaya) yang sudah menjadi keluarga walaupun hanya sebentar dipertemukan. Terima kasih untuk semua kenangan yang diberikan. Sukses untuk kita semua.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, sehingga kritik dan saran sangat diharapkan. Semoga tesis ini dapat memberikan manfaat guna memperluas wawasan keilmuan pembacanya.

Surabaya, Januari 2018

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR TABEL	xv
DAFTAR GAMBAR	xvii
DAFTAR LAMPIRAN	xix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	6
1.3 Tujuan Penelitian.....	7
1.4 Manfaat Penelitian.....	7
1.5 Batasan Masalah Penelitian	7
BAB 2 TINJAUAN PUSTAKA	9
2.1 Analisis Survival	9
2.1.1 Fungsi <i>Survival</i> dan Fungsi <i>Hazard</i>	11
2.1.2 Kurva <i>Survival</i> Kaplan-Meier dan Log-Rank	13
2.2 <i>Cox Proportional Hazard</i>	15
2.2.1 Asumsi <i>Proportional Hazard</i>	16
2.3 <i>Support Vector Machine</i> (SVM).....	19
2.4 <i>Least Squares Support Vector Machine</i> (LS-SVM).....	23
2.5 <i>Survival Support Vector Machine</i> (SURSVM)	24
2.6 <i>Survival Least Squares Support Vector Machine</i> (SURLSSVM)	25

2.7	Kriteria Keباikan Model	27
2.8	Pembangkitan Data <i>Survival Time</i>	29
2.9	Metode <i>Feature Selection (Backward Elimination)</i>	31
2.10	Kanker Serviks	32
BAB 3	METODE PENELITIAN	37
3.1	Kajian Teoritis Model SURLS-SVM	37
3.2	Aplikasi Model SURLS-SVM	38
3.2.1	Data Simulasi	38
3.2.2	Kerangka Konsep untuk Data Riil	40
3.2.3	Sumber Data dan Variabel Penelitian	43
3.2.4	Langkah Analisis Data	45
3.2.5	Struktur Data Penelitian	49
BAB 4	HASIL DAN PEMBAHASAN	51
4.1	Kajian Teoritis Model SURLS-SVM	51
4.2	Penerapan SURLS-SVM dan Cox PHM untuk Data Simulasi	58
4.2.1	Performansi SURLS-SVM dan Cox PHM pada Data Linier dan Non Linier	59
4.2.2	Penerapan <i>Feature Selection</i> pada Data Simulasi	65
4.3	Analisis Data Pasien Kanker Serviks	68
4.3.1	Karakteristik Pasien Kanker Serviks	68
4.3.2	Analisis Kurva <i>Survival</i> Kaplan Meier dan Uji Log Rank	75
4.4	Analisis Data Pasien Kanker Serviks dengan Cox PHM dan SURLS-SVM	85
4.4.1	Uji Asumsi Proporsional Hazard Pada Setiap Variabel	85
4.4.2	Fungsi <i>Survival</i> Kumulatif dan Fungsi Hazard Kumulatif	86
4.4.3	Penerapan Model Cox PHM pada Data Pasien Kanker Serviks	87
4.4.4	Penerapan Model SURLS-SVM pada Data Pasien Kanker Serviks	89

4.4.5 Penerapan <i>Feature Selection</i> pada Data Pasien Kanker Serviks	90
BAB 5 KESIMPULAN DAN SARAN	93
5.1 Kesimpulan	93
5.2 Saran	94
DAFTAR PUSTAKA	95
LAMPIRAN	99
BIOGRAFI PENULIS	135

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 2.1 <i>Staging</i> Kanker Serviks.....	32
Tabel 3.1 Distribusi Variabel Prediktor	39
Tabel 3.2 Variabel Prediktor yang Diinteraksikan.....	40
Tabel 3.3 Variabel Respon.....	43
Tabel 3.4 Variabel Prediktor Pasien Kanker Serviks.....	44
Tabel 3.5 Struktur Data <i>Survival</i>	49
Tabel 4.1 Contoh Data <i>Survival</i>	54
Tabel 4.2 Statistika Deskriptif Faktor Usia, Usia Pertama Menikah, dan Usia Menstruasi Pertama	69
Tabel 4.3 Statistika Deskriptif Faktor Siklus Mentruasi, Lama Menstruasi, dan Paritas	70
Tabel 4.4 Tabulasi Silang antara Status Pasien dengan Status Komplikasi.	71
Tabel 4.5 Tabulasi Silang antara Status Pasien dengan Status Anemia.....	71
Tabel 4.6 Tabulasi Silang antara Status Pasien dengan Jenis Pengobatan ..	72
Tabel 4.7 Tabulasi Silang antara Status Pasien dengan Stadium.....	73
Tabel 4.8 Tabulasi Silang antara Status Pasien dengan Jenis KB	74
Tabel 4.9 Tabulasi Silang antara Status Pasien dengan Tingkat Pendidikan	75
Tabel 4.10 Uji Asosiasi Variabel	75
Tabel 4.11 Hasil Uji Log-Rank per Variabel	84
Tabel 4.12 Hasil Uji Asumsi PH Per Variabel.....	85
Tabel 4.13 Hasil Estimasi Parameter Model Cox PHM	87
Tabel 4.14 Performansi Model Cox PHM pada Data Pasien Kanker Serviks	88
Tabel 4.15 Performansi Model SURLS-SVM pada Data Pasien Kanker Serviks	89
Tabel 4.16 Performansi Cox PHM dan SURLS-SVM Setelah <i>Feature</i> <i>Selection</i>	91
Tabel 4.17 Pengaruh Variabel Terpilih Terhadap Nilai c-index.....	91

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Kurva <i>Survival</i>	11
Gambar 2.2 Kurva Fungsi <i>Hazard</i>	12
Gambar 2.3 Kurva Kaplan-Meier	14
Gambar 2.4 Grafik $-\ln(-\ln S(t))$	16
Gambar 2.5 Grafik Observasi-Prediksi	17
Gambar 2.6 SVM dengan <i>Classifier</i> Linear.....	19
Gambar 2.7 Non-Linear SVM.....	21
Gambar 2.8 Ilustrasi Penghitungan c-index	28
Gambar 2.9 Tahapan <i>Backward Elimination</i>	31
Gambar 3.1 Kerangka Konsep Ketahanan Hidup Pasien Kanker Serviks..	42
Gambar 3.2 Langkah Analisis	48
Gambar 4.1 Performansi Data Simulasi Linier ($n=1000, p=17$).....	59
Gambar 4.2 Performansi Data Simulasi Non-Linier ($n=1000, p=17$).....	60
Gambar 4.3 Persentase Signifikansi Variabel	66
Gambar 4.4 Kenaikan Nilai c-index Setelah <i>Feature Selection</i>	67
Gambar 4.5 Persentase Pasien Kanker Serviks Berdasarkan Status Tersensor	68
Gambar 4.6 Kurva Kaplan Meier Pasien Kanker Serviks.....	76
Gambar 4.7 Kurva Kaplan Meier Berdasarkan Usia.....	76
Gambar 4.8 Kurva Kaplan Meier Berdasarkan Komplikasi	77
Gambar 4.9 Kurva Kaplan Meier Berdasarkan Anemia	78
Gambar 4.10 Kurva Kaplan Meier Berdasarkan Jenis Pengobatan	78
Gambar 4.11 Kurva Kaplan Meier Berdasarkan Stadium.....	79
Gambar 4.12 Kurva Kaplan Meier Berdasarkan Usia Saat Menikah Pertama	80
Gambar 4.13 Kurva Kaplan Meier Berdasarkan Usia Saat Menstruasi Pertama	80
Gambar 4.14 Kurva Kaplan Meier Berdasarkan Siklus Menstruasi	81
Gambar 4.15 Kurva Kaplan Meier Berdasarkan Lama Menstruasi	82

Gambar 4.16 Kurva Kaplan Meier Berdasarkan Paritas.....	82
Gambar 4.17 Kurva Kaplan Meier Berdasarkan Jenis KB	83
Gambar 4.18 Kurva Kaplan Meier Berdasarkan Tingkat Pendidikan	84
Gambar 4.19 Kurva <i>Survival</i> Kumulatif.....	86
Gambar 4.20 Kurva <i>Hazard</i> Kumulatif	86

DAFTAR LAMPIRAN

	Halaman
Lampiran 1	Sertifikat Laik Etik 99
Lampiran 2	Data <i>Survival</i> Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya 100
Lampiran 3	Tabulasi Silang Status Tersensor dengan Variabel Stadium dan Tingkat Pendidikan 101
Lampiran 4	Ukuran Performansi Berdasarkan Persentase Tersensor untuk Data Linier dengan <i>Sample Size</i> 100..... 102
Lampiran 5	Ukuran Performansi Berdasarkan Persentase Tersensor untuk Data Linier dengan <i>Sample Size</i> 1000..... 103
Lampiran 6	Ukuran performansi Berdasarkan Persentase Tersensor untuk Data Linier dengan <i>Sample Size</i> 5000..... 104
Lampiran 7	Ukuran Performansi Berdasarkan Persentase Tersensor untuk Data Non-Linier dengan <i>Sample Size</i> 100..... 105
Lampiran 8	Ukuran Performansi Berdasarkan Persentase Tersensor untuk Data Non-Linier dengan <i>Sample Size</i> 1000..... 106
Lampiran 9	Ukuran Performansi Berdasarkan Persentase Tersensor untuk Data Non-Linier dengan <i>Sample Size</i> 5000..... 107
Lampiran 10	Iterasi Parameter untuk Model SURLS-SVM pada Data Pasien Kanker Serviks Berdasarkan Nilai C-Index..... 108
Lampiran 11	Syntax R untuk Kurva Kaplan Meier dan Uji Log-Rank.... 109
Lampiran 12	Syntax Program R untuk Menguji Asumsi Proportional Hazard dengan <i>Goodness of Fit</i> 114
Lampiran 13	Syntax R untuk Kurva Fungsi <i>Survival</i> dan Kurva Fungsi Hazard..... 114
Lampiran 14	Syntax R untuk Membangkitkan Data dengan Pendekatan Parametrik..... 115
Lampiran 15	Syntax R untuk Function yang Digunakan pada Model SURLS-SVM..... 116

Lampiran 16	Syntax R untuk Analisis dengan Cox PHM dan SURLS-SVM serta menentukan Tuning Parameter Optimal dengan Metode <i>Grid Search</i>	119
Lampiran 17	Syntax R untuk <i>Feature Selection</i> dengan Metode <i>Backward Elimination</i> pada model Cox PHM dan SURLS-SVM.....	122

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Kemajuan ilmu pengetahuan dan teknologi memunculkan berbagai macam metode statistika yang dapat diterapkan dalam berbagai bidang, salah satunya adalah bidang kedokteran. Metode statistika dalam bidang kedokteran salah satunya digunakan untuk menganalisa waktu bertahannya seorang pasien terhadap suatu penyakit tertentu atau yang dikenal dengan istilah *survival time*. *Survival time* ini selanjutnya dianalisa dengan menggunakan analisis *survival*. Analisis *survival* merupakan salah satu dari metode statistika yang bertujuan untuk mempelajari dan memodelkan hubungan antara faktor-faktor yang beresiko mempengaruhi kelangsungan hidup pasien dan waktu terjadinya kematian pasien (Abadi, 2011).

Menurut Kleinbaum dan Klein (2012) pada analisis *survival* terdapat dua macam fungsi yang diamati, yaitu fungsi *survival* dan fungsi *hazard*. Fungsi *survival* menyatakan probabilitas pasien tidak mengalami *event (failure)*, sedangkan fungsi *hazard* menyatakan laju terjadinya *event*. Untuk memodelkan *failure* yang terjadi dapat digunakan *hazard* model. Analisis *survival* dapat menggunakan pendekatan parametrik, semi-parametrik, dan non-parametrik. Pada awalnya pendekatan yang digunakan dalam analisis *survival* adalah pendekatan parametrik. Dalam pendekatan parametrik, *baseline hazard* terdefinisi sesuai dengan distribusi dari data. Terdapat batasan distribusi yang dapat digunakan dalam *survival time* yaitu bersifat non-negatif (Kleinbaum dan Klein, 2012), beberapa diantaranya adalah Weibull dan Eksponensial. Pendekatan yang digunakan untuk membentuk fungsi *hazard* antara lain probit dan logit. Penelitian dengan pendekatan parametrik sudah pernah dilakukan salah satunya oleh Lestari (2016), dengan objek penelitian adalah pasien penderita jantung koroner dan yang diamati adalah laju perbaikan klinis menggunakan *multi period logit*. Penelitian lain dilakukan oleh Furqon (2013) yang menggunakan regresi Weibull untuk mengetahui faktor-faktor yang mempengaruhi laju perbaikan klinis penderita stroke. Analisis *survival* dengan pendekatan parametrik lebih sederhana karena

baseline hazard dapat didefinisikan secara jelas sesuai dengan distribusi dari *survival time*, namun juga memiliki kelemahan yaitu harus diketahui distribusi dari *survival time* agar dapat dimodelkan dengan tepat (Kleinbaum dan Klein, 2012).

Pendekatan lain yang dapat digunakan untuk analisis *survival* adalah pendekatan semi-parametrik. Model untuk semi-parametrik yang paling populer adalah model Cox Proporsional Hazard Model (Cox PHM) atau dikenal dengan regresi Cox yang pertama kali diperkenalkan oleh Cox pada 1972. Banyak penelitian *survival* yang menggunakan Cox PHM. Wijaya (2015) pada penelitiannya membahas mengenai laju perbaikan klinis pasien penyakit sindrom koroner akut (SKA) dengan model Cox PHM. Penelitian lain dilakukan oleh Afifah dan Purnami (2016) yang menguji asumsi *proportional hazard* pada data penderita kanker serviks, dan dari penelitian tersebut diketahui bahwa variabel stadium 4 tidak memenuhi asumsi PH, artinya penderita kanker serviks stadium 4 memiliki peluang ketahanan hidup yang berubah-ubah setiap waktunya. Purnami, Inayati, Sari, Chosuvivatwong, dan Sriplung (2015) meneliti mengenai kanker serviks dengan menggunakan Regresi Stratified Cox yang merupakan pengembangan Cox PHM karena terdapat variabel stadium yang melanggar asumsi PH. Pendekatan semi-parametrik pada model Cox pada penelitian-penelitian yang telah dilakukan memiliki kelebihan karena tidak membutuhkan asumsi suatu distribusi pada *baseline hazard* sehingga *survival time* tidak perlu diketahui distribusinya. Namun, model Cox PHM juga memiliki kekurangan karena variabel yang diamati harus memenuhi asumsi proporsioanl dan kurang bisa diaplikasikan pada data *survival* yang berukuran besar atau bersifat *high dimensional data* (Van Belle, Pelckmans, Suykens, dan Vam Huffel, 2011).

Seiring dengan berkembangnya ilmu pengetahuan, muncul pengembangan baru pada analisis data *survival* yaitu pendekatan non-parametrik. Pendekatan ini menyempurnakan kekurangan pada dua pendekatan sebelumnya karena tidak memerlukan asumsi proporsional dan dapat digunakan untuk data *survival* yang jumlahnya besar. Metode yang menggunakan pendekatan non-parametrik salah satunya adalah *Multivariate Adaptive Regression Splines* (MARS) yang dapat digunakan untuk analisis *survival*. Penelitian yang dilakukan oleh Nisa dan Budiantara (2013) menggunakan MARS untuk analisis *survival* pada kasus DBD.

Penelitian lain dilakukan oleh Van Belle, Pelckmans, Suykens, dan Vam Huffel (2010a) yang menganalisis data *survival* menggunakan Partial Logistic Artificial Neural Network (PLANN) untuk data kanker payudara. Dalam penelitian tersebut, Van Belle *et al.* (2010a) juga menyebutkan bahwa kekurangan PLANN adalah data perlu direplikasi pada tiap interval waktu selama objek tersebut belum mengalami *event*, selain itu PLANN memberikan hasil yang kurang akurat untuk memprediksi koefisien variabel prediktor pada data yang berdimensi tinggi. Semakin pesatnya perkembangan ilmu pengetahuan, Vapnik (1998) memperkenalkan metode SVM dan menyebutkan bahwa metode SVM adalah metode *machine learning* yang berhasil melakukan prediksi, baik untuk klasifikasi dan regresi. SVM yang banyak dikenal sebelumnya hanya digunakan untuk klasifikasi dimana data yang diklasifikasikan merupakan data yang utuh. Sedangkan pada *Survival SVM* (SUR-SVM) data yang tersedia adalah data tersensor, sehingga diperlukan suatu reformulasi untuk fungsi tujuan maupun kendala. *Survival SVM* dapat digunakan untuk memprediksi *ranking survival time* pasien sehingga dapat diketahui pasien mana yang memiliki peluang bertahan hidup lebih besar. Dalam perkembangannya, *survival SVM* dapat digunakan untuk memprediksi *survival time* dengan pendekatan ranking. Van Belle, Pelckmans, Suykens, dan Vam Huffel (2010b) telah melakukan penelitian mengenai pengembangan metode SUR-SVM. Metode ini bekerja dengan *quadratic programming* yang mampu menghasilkan solusi optimal dalam menentukan fungsi *Lagrange*. Namun, hal ini kurang efisien jika diterapkan pada data berdimensi tinggi (Khaulasari, 2016). Oleh karena itu dikembangkan *Least Squares Support Vector Machine* (LS-SVM) yang pertama kali diperkenalkan oleh Suykens dan Vandewalle (1999).

LS-SVM menggunakan fungsi kendala (*constrain*) berupa persamaan sehingga solusinya dapat diperoleh dengan persamaan linier. Hal ini berbeda dengan SVM, dimana fungsi kendalanya berupa pertidaksamaan sehingga diselesaikan dengan *quadratic programming*. Suykens dan Vandewalle (1999) menyatakan bahwa LS-SVM memberikan hasil yang lebih baik dibandingkan dengan SVM dalam hal proses perhitungan, konvergensi yang diperoleh lebih cepat, dan presisi bagus. LS-SVM ini juga dikembangkan untuk data *survival*, salah satunya adalah penelitian dari Van Belle *et al.* (2010a) yang menerapkan LS-SVM

untuk data kanker payudara di Jerman. Konsep dari *Survival LS-SVM* (SURLS-SVM) ini adalah dengan meranking pasien berdasarkan waktu *failure* dan waktu tersensor. Dalam penelitiannya, Van Belle *et al.* (2010a) memperkenalkan *prognostic index* yang merupakan pengganti dari hazard function pada Cox PHM. *Prognostic index* yang diperoleh selanjutnya digunakan untuk mengukur kebaikan model melalui concordance index (c-index), log-rank, dan hazard ratio.

Model SURLS-SVM dalam penerapannya tidak dapat digunakan secara langsung untuk mengetahui variabel prediktor yang berpengaruh secara signifikan terhadap ukuran kebaikan model (Van Belle *et al.*, 2011). Untuk mengetahui hal itu, dapat dilakukan dengan menggunakan *feature selection* dimana dalam hal ini dilakukan dengan metode *backward selection*. Salah satu penelitian yang menjelaskan tentang *feature selection* pada SVM telah digunakan oleh Goli, *et.al* (2016) yang menerapkannya pada pasien kanker payudara dimana hasilnya adalah: (i) bila semua *features* disertakan dalam model, SVR dan Cox PHM memiliki performansi yang sama dan (ii) ketika *feature selection* diterapkan pada SVR, diperoleh c-index SVR yang lebih baik dari yang dihasilkan Cox PHM.

Penerapan model-model yang telah diuraikan di atas dalam bidang kesehatan banyak digunakan untuk memprediksi waktu bertahan hidup pasien kanker. Salah satu jenis kanker yang menjadi perhatian adalah kanker serviks. Berdasarkan WHO (2005) kanker serviks merupakan kanker yang paling banyak diderita oleh perempuan setelah kanker payudara, dimana perkiraan kasusnya pada tahun 2000 adalah 471.000 kasus baru dan 233.000 diantaranya meninggal dunia. Dari data tersebut, 80% penderita berasal dari negara berkembang di Asia Selatan, Asia Tenggara, Sub-Saharan Afrika, Amerika Tengah, dan Amerika Selatan. Berdasarkan catatan dari Kementerian Kesehatan RI pada tahun 2013 terdapat 98.692 kasus kanker serviks di Indonesia. Dalam penelitian Ocviyanti dan Handoko (2013) menyebutkan bahwa kanker serviks dari segi jumlah penderita adalah tertinggi kedua setelah kanker payudara tetapi menjadi yang paling tinggi jika dilihat dari segi penyebab kematian, dan pada tahun 2025 diperkirakan kasus barukanker serviks di Indonesia meningkat 74%.

Hingga saat ini sudah banyak penelitian yang menyebutkan bahwa penyebab kanker serviks adalah infeksi *Human Papilloma Virus* (HPV) yang

onkogenik. Menurut Howley dan Lowy (2001) jenis HPV yang bersifat onkogenik adalah tipe 16, 18, 45, 31, 33, 52, 58, 35, 59, 56, 51, 39, 68, 73, dan 82. Penyebab utama kanker serviks adalah tipe 16 dan 18 (sekitar 70%). HPV menginfeksi penderita kebanyakan melalui kontak seksual. Inveksi virus ini, umumnya tidak memberikan gejala apapun pada penderitanya padahal saat itu penderita sudah terinfeksi HPV, bahkan ketika infeksi ini sudah menyebabkan lesi prakanker yaitu perubahan sel pada lapisan epitel serviks yang berpotensi menjadi kanker serviks. Lesi prakanker ini dapat dideteksi dengan tes Pap (*papsmear*) dan dapat juga menggunakan *Hybrid Capture HPV Test* (Ocviyanti dan Handoko, 2013).

Selain dari sudut pandang medis, penelitian yang membahas mengenai kanker serviks dari sudut pandang statistika sudah banyak dilakukan. Inayati dan Purnami (2015) melakukan penelitian mengenai ketahanan hidup pasien kanker serviks dengan menggunakan pendekatan non-parametrik yaitu Kaplan Meier dan Uji Log Rank. Penelitian tersebut menggunakan enam faktor yaitu usia, stadium, jenis pengobatan, penyakit penyerta, komplikasi, dan status anemia. Berdasarkan hasil dari kurva Kaplan-Meier, variabel yang terdapat perbedaan kurva *survival* adalah stadium, jenis pengobatan, penyakit penyerta, dan komplikasi. Lalu berdasarkan uji log-rank, variabel yang menunjukkan perbedaan kurva *survival* adalah stadium, jenis pengobatan, dan komplikasi. Penelitian analisis *survival* dengan metode nonparametrik juga dilakukan oleh Respita (2017), yang dalam penelitiannya menganalisis data *survival* pasien kanker serviks yang menjalani rawat inap di RS dr.Soetomo pada bulan Januari 2014 sampai dengan Desember 2014 dengan model regresi *splines* dan model MARS. Variabel yang digunakan dalam penelitian tersebut adalah usia, stadium, jenis pengobatan kemoterapi, jenis pengobatan transfusi PRC, jenis pengobatan operasi, penyakit penyerta, komplikasi dan status anemia. Berdasarkan model yang digunakan, variabel yang berpengaruh pada tingkat ketahanan hidup pasien kanker serviks adalah usia , stadium, jenis pengobatan transfusi PRC dan penyakit penyerta.

Pada Cox PHM, model yang terbentuk dinyatakan dalam hazard function karena prediksi *survival time* tidak bisa diperoleh. Karena *survival time* yang tidak dapat diprediksi maka muncul *prognostic index*. Prognostic dalam bidang kedokteran diartikan sebagai perkiraan hasil penyembuhan pasien berdasarkan pada

treatment atau perawatan medis yang telah dilakukan atau dapat juga dilihat berdasarkan faktor intrinsik dari penyakit yang dialami. *Prognostic index* inilah yang digunakan untuk menentukan kebaikan model, dengan menggunakan ukuran c-index, hasil dari log-rank test, dan hazard ratio. Prediksi *prognostic index* ditentukan dengan menggunakan dua model yaitu Cox PHM dan SURLS-SVM. Salah satu hal yang menjadi perhatian dalam mengukur kebaikan model adalah bagaimana keterurutan atau kesesuaian antara *survival time* sebenarnya dengan *prognostic index* yang merupakan hasil dari prediksi. Dalam penelitian ini data yang digunakan adalah data pasien kanker serviks di RS dr. Soetomo Surabaya yang dihitung nilai *prognostic index* menggunakan dua model dan dibandingkan model manakah yang lebih baik dengan menggunakan tiga ukuran kebaikan model yang telah disebutkan sebelumnya. Sebagai pembandingan, juga akan dilakukan studi simulasi dengan menggunakan metode yang sama pada berbagai ukuran data.

1.2 Perumusan Masalah

Data *survival* tidak selalu mempunyai distribusi parametrik, dan terdapat data yang tersensor sehingga untuk melakukan analisis diperlukan suatu metode yang dapat diterapkan untuk data non-parametrik dan mampu mengakomodasi data tersensor. Pada Cox PHM model yang terbentuk dituliskan dalam hazard function karena *survival time* prediksi tidak dapat diperoleh, oleh karena itu digunakan *prognostic index* yang menggantikan hazard function pada Cox PHM. Berdasarkan uraian tersebut, maka permasalahan yang dibahas dalam penelitian ini adalah bagaimana mendapatkan persamaan prognostic-index untuk *Survival Least Square-SVM* (SURLS-SVM). Selanjutnya bagaimana aplikasi persamaan *prognostic index* yang telah diperoleh untuk data pasien kanker serviks dan data simulasi serta manakah variabel yang secara signifikan mempengaruhi performansi model SURLS-SVM.

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah di atas, tujuan dari penelitian ini adalah sebagai berikut:

1. Melakukan kajian dalam menentukan *prognostic index* pada SURLS-SVM.
2. Menerapkan model SURLS-SVM untuk analisa data simulasi dan data *survival* pasien kanker serviks di RSUD dr. Soetomo Surabaya.
3. Membandingkan performansi SURLSSVM dengan regresi Cox PHM.
4. Mendapatkan variabel prediktor yang berpengaruh secara signifikan terhadap performansi model SURLS-SVM (dalam hal ini berdasarkan nilai c-index) menggunakan metode *backward selection*.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah memberikan alternatif metode yang dapat digunakan untuk analisis data *survival* dengan ukuran yang sangat besar atau memiliki variabel prediktor yang banyak, dimana dalam penelitian ini model non-parametrik yang digunakan adalah SURLS-SVM. Selain itu hasil dari penelitian ini diharapkan dapat memberikan informasi tambahan bagi RS dr. Soetomo mengenai laju kematian pasien kanker serviks sehingga dapat digunakan sebagai salah satu pertimbangan dalam pemberian pengobatan yang sesuai berdasarkan prediksi pasien yang masuk dalam *high risk* maupun *low risk*.

1.5 Batasan Masalah Penelitian

Dalam penelitian ini, ruang lingkup permasalahan dibatasi pada beberapa hal sebagai berikut:

1. Data yang digunakan sebagai data riil adalah data pasien rawat inap pasien kanker serviks di RS dr. Soetomo Surabaya pada tahun 2014-2016.
2. Kernel yang digunakan untuk model SURLS-SVM adalah *Radial Basis Function*.
3. Tipe data tersensor yang digunakan adalah data yang tersensor kanan.
4. *Event* atau kejadian dalam penelitian ini adalah pasien meninggal.
5. *Survival time* pada studi simulasi tidak dikaitkan dengan distribusi *survival time* pada data riil yang digunakan sebagai studi kasus.

(Halaman ini sengaja dikosongkan)

BAB 2

TINJAUAN PUSTAKA

2.1 Analisis *Survival*

Analisis *survival* adalah suatu metode statistik dimana variabel yang menjadi perhatian adalah waktu hingga suatu kejadian (*event*) terjadi atau disebut dengan waktu *survival*. Waktu yang digunakan dalam analisis *survival* bisa dalam satuan hari, minggu, bulan atau bahkan tahun yang diperoleh dari awal pengamatan hingga *event* terjadi, sedangkan kejadian (*event*) dapat berupa *event* positif dan *event* negatif (Kleinbaum dan Klein, 2012). Dalam dunia kesehatan, *event* positif identik dengan sembuh atau kondisi membaik dari suatu penyakit, sedangkan *event* negatif identik dengan meninggal atau kambuh.

Dalam menentukan waktu *survival* harus memperhatikan beberapa hal, yaitu (Cox, 1972):

1. waktu awal atau dimulainya suatu kejadian (*time origin/starting point*),
2. definisi *event* dari seluruh kejadian yang diamati harus jelas,
3. skala pengukuran yang digunakan.

Pada suatu pengamatan tidak jarang ditemukan ada beberapa objek yang waktu *survival*nya tidak diketahui secara pasti, sehingga menyebabkan data dari objek tersebut tersensor (*censored data*). Penyebab terjadinya *censored data* adalah sebagai berikut (Kleinbaum dan Klein, 2012):

1. *termination of study*, yaitu objek yang diobservasi belum mencapai *event* sedangkan masa penelitian sudah berakhir,
2. *lost to follow-up*, yaitu objek tidak mengikuti *treatment* yang dianjurkan sehingga pengamatan tidak dapat dilakukan secara keseluruhan, misalnya pindah atau tidak bersedia diberikan *treatment*,
3. *withdraws from the study*, yaitu *treatment* yang diberikan dihentikan karena alasan tertentu, misalnya memberikan efek samping terhadap pasien sehingga kesehatannya memburuk.

Menurut Lee dan Wang (2003) terdapat beberapa cara pengklasifikasian tipe data tersensor, yaitu:

1. tersensor tipe I

Terjadi jika semua nilai c_i sama atau $c_i = c$ dan nilai konstan tersebut ditentukan oleh peneliti. Atau dengan kata lain periode penelitian telah ditentukan dan objek yang akan diteliti masuk ke dalam penelitian pada waktu yang sama,

2. tersensor tipe II

Terjadi jika pengamatan berhenti setelah d pengamatan, dimana banyaknya pengamatan yang mengalami *event (failure)* sudah ditentukan di awal oleh peneliti. Pada data tersensor tipe II, objek masuk dalam penelitian pada waktu yang sama dan penelitian berhenti jika jumlah objek yang mengalami *event* sudah sesuai dengan yang ditentukan di awal penelitian,

3. tersensor tipe III

Data tersensor tipe III disebut juga tersensor secara progresif terjadi jika setiap objek masuk dalam penelitian pada waktu yang berbeda-beda selama periode penelitian. Data seperti ini sering terjadi pada penelitian klinis.

Selain klasifikasi di atas, menurut Collett (2003) ada jenis pengklasifikasian lain untuk data tersensor yaitu sebagai berikut:

1. sensor kanan

Jika objek yang diamati tidak mengalami *event* sampai penelitian selesai. Selain itu sensor kanan juga dapat terjadi karena objek hilang dari pengamatan (*lost to follow-up*) atau objek mengalami *event* tetapi tidak disebabkan oleh faktor utama yang diteliti (*withdrawn*) saat penelitian masih berlangsung. Dengan kata lain, data tersensor kanan terjadi jika waktu *survival* suatu objek tidak lengkap sampai dengan penelitian selesai,

2. sensor kiri

Jika objek mengalami *event* sebelum data mulai dikumpulkan. Dengan kata lain, data tersensor kiri terjadi jika waktu *survival* seorang individu yang sebenarnya kurang dari atau sama dengan waktu *survival* individu saat dilakukan penelitian. Atau dapat dikatakan bahwa saat penelitian dilakukan, individu sudah mengalami *event* tetapi tidak diketahui secara pasti kapan *event* itu terjadi,

3. sensor interval

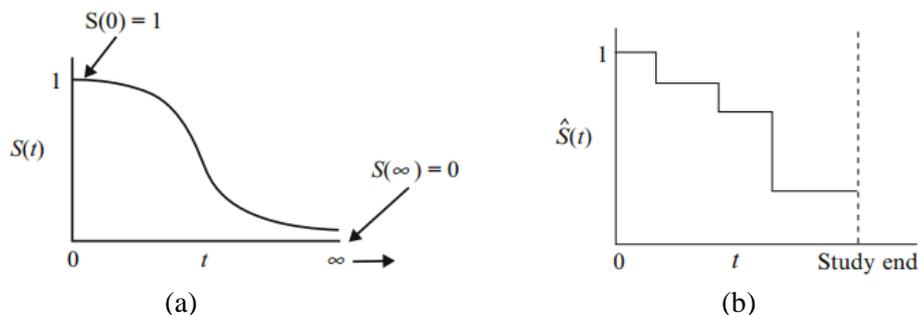
Jika objek mengalami *event* diantara waktu t_1 dan t_2 tetapi tidak diketahui dengan pasti kapan *event* terjadi.

Analisis *survival* yang dilakukan memiliki beberapa tujuan umum, yaitu:

1. mengestimasi dan menginterpretasikan fungsi *survival* dan fungsi *hazard* dari data *survival*,
2. membandingkan fungsi *survival* dan fungsi *hazard*,
3. mengetahui hubungan variabel prediktor dengan waktu *survival*.

2.1.1 Fungsi *Survival* dan Fungsi *Hazard*

Dalam analisis *survival* terdapat dua macam fungsi yang sering digunakan yaitu fungsi *survival* dan fungsi *hazard*. Pada sub-bab ini semua teori diambil dari Kleinbaum dan Klein (2012). Fungsi *survival* dinotasikan $S(t)$, merupakan probabilitas suatu objek dapat bertahan atau tidak mengalami *event (failure)* sampai pada waktu tertentu (Kleinbaum dan Klein, 2012). Fungsi *survival* dapat digambarkan dalam kurva berikut:



Gambar 2.1 (a) Kurva Fungsi *Survival* dalam teori

(b) Kurva Fungsi *Survival* dalam empiris

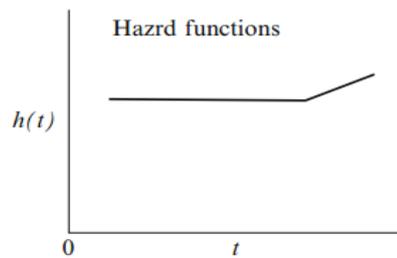
Fungsi *survival* jika dinyatakan dalam distribusi kumulatif adalah sebagai berikut :

$$S(t) = P(T > t) = \int_t^{\infty} f(u) du \quad (2.1)$$

dengan T adalah waktu yang dibutuhkan objek sampai mengalami *event (failure)* atau dikenal dengan *time to failure*. Karena $S(t)$ adalah suatu probabilitas maka nilainya berada pada interval 0 sampai dengan 1. Berdasarkan Gambar 2.1 (a) dan (b), saat awal pengamatan probabilitas total ketahanan pasien adalah 1 karena pada

saat awal pengamatan belum ada pasien yang mengalami *event*, dan semakin panjang waktu pengamatan semakin berkurang hingga mendekati 0.

Fungsi *hazard* dinotasikan dengan $h(t)$ merupakan laju terjadinya *event* (*failure*) suatu objek. Fungsi *hazard* bukan merupakan suatu probabilitas, tetapi merupakan fungsi non-negatif yang tidak memiliki batas atas (*upper bound*). Kurva dari fungsi *hazard* adalah sebagai berikut:



Gambar 2.2 Kurva Fungsi *Hazard*

Dari Gambar 2.2 dapat dilihat bahwa dengan bertambahnya waktu (t), laju seorang pasien mengalami *event* semakin cepat. Hal ini berlawanan dengan nilai dari fungsi *survival*. Secara matematis fungsi *hazard* dapat dinyatakan dalam persamaan berikut:

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

Berdasarkan definisi fungsi *hazard* tersebut dapat diketahui hubungan antara fungsi *hazard* dan fungsi *survival* dengan menggunakan teori probabilitas bersyarat dan fungsi distribusi dari T . Diketahui probabilitas bersyarat

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ dengan } P(A \cap B) \text{ merupakan probabilitas kejadian bersama}$$

antara A dan B . Sedangkan definisi probabilitas bersyarat menurut fungsi *hazard* adalah sebagai berikut :

$$\frac{P(t \leq T < t + \Delta t)}{P(T > t)} = \frac{F(t + \Delta t) - F(t)}{S(t)} \quad (2.3)$$

dengan $F(t)$ adalah fungsi distribusi dari T , sehingga diperoleh :

$$h(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \frac{1}{S(t)} \quad (2.4)$$

dengan

$$F'(t) = f(t) = \lim_{\Delta t \rightarrow 0} \left\{ \frac{F(t + \Delta t) - F(t)}{\Delta t} \right\} \quad (2.5)$$

merupakan derivatif dari $F(t)$. Sehingga diperoleh hubungan antara fungsi *survival* dan fungsi *hazard* adalah sebagai berikut :

$$h(t) = \frac{f(t)}{S(t)} \quad (2.6)$$

Selain persamaan matematis fungsi *survival* di atas, terdapat penurunan persamaan lain dari fungsi *survival* yaitu $F(t) = 1 - S(t)$ sehingga diperoleh

$\int_0^t f(u) du = 1 - S(t)$, jika kedua sisi diturunkan terhadap t maka akan diperoleh

$f(t) = \frac{d(1 - S(t))}{dt}$, dan persamaan dari fungsi *hazard* dapat diturunkan menjadi

persamaan berikut

$$h(t) = \frac{\frac{d(1 - S(t))}{dt}}{S(t)} = -\frac{1}{S(t)} \frac{d(S(t))}{dt} \quad (2.7)$$

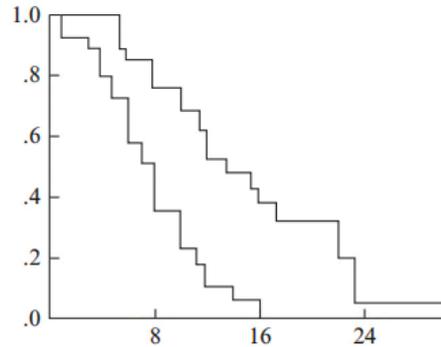
Dengan mengintegalkan kedua sisi persamaan (2.7) akan diperoleh persamaan berikut

$$-\int_0^t h(u) du = \int_0^t \frac{1}{S(u)} d(S(u)) = \ln(S(t)) \quad (2.8)$$

Fungsi *hazard* kumulatif dinyatakan dengan $H(t) = \int_0^t h(u) du$ atau $-H(t) = \ln S(t)$

2.1.2 Kurva Kaplan-Meier dan Log-Rank

Kurva Kaplan-Meier merupakan kurva berbentuk fungsi tangga yang menggambarkan hubungan antara estimasi fungsi *survival* pasien yang memiliki nilai antara nol sampai dengan satu pada waktu t (terletak pada sumbu vertikal) dengan *survival time* (terletak pada sumbu horisontal). Berikut adalah contoh dari kurva Kaplan-Meier pada kasus pemberian *placebo* dan *treatment* pada pasien Leukimia.



Gambar 2.3 Kurva Kaplan-Meier

Persamaan dari estimasi fungsi *survival* adalah sebagai berikut:

$$\begin{aligned} \hat{S}(t_f) &= \prod_{i=1}^n \hat{P} [T > t_i | T \geq t_i] \\ &= \prod_{i=1}^{n-1} \hat{P} [T > t_i | T \geq t_i] \times \hat{P} [T > t_n | T \geq t_n] \\ \hat{S}(t_f) &= \hat{S}(t_{n-1}) \times \hat{P} [T > t_n | T \geq t_n] \end{aligned} \quad (2.9)$$

Kurva Kaplan-Meier juga dapat digunakan untuk menggambarkan hubungan antara estimasi fungsi *survival* dengan *survival time* sesuai dengan kelompok faktor yang diamati. Namun juga diperlukan uji selanjutnya menggunakan uji Log-Rank. Uji Log-Rank digunakan untuk membandingkan kurva *survival* dalam kelompok yang berbeda apakah terdapat perbedaan kurva *survival* atau tidak.

Hipotesis yang digunakan untuk uji Log-Rank adalah sebagai berikut:

H_0 : Tidak ada perbedaan kurva *survival*

H_1 : Minimal terdapat satu perbedaan kurva *survival*

Dengan statistik uji sebagai berikut:

$$\chi^2 \approx \sum_{g=1}^G \frac{(O_g - E_g)^2}{E_g} \quad (2.10)$$

dengan :

$$O_g - E_g = \sum_{i=1}^n (m_{gi} - e_{gi}) \quad e_{gi} = \left(\frac{n_{gi}}{\sum_{g=1}^G \sum_{i=1}^n n_{gi}} \right) \left(\sum_{g=1}^G \sum_{i=1}^n m_{gi} \right)$$

Keterangan :

O_g = nilai observasi individu pada grup ke-g

E_g = nilai ekspektasi individu pada grup ke-g

m_{gi} = jumlah individu pada grup ke-g yang mengalami *event* pada waktu t_i

n_{gi} = jumlah individu yang beresiko mengalami *event* seketika pada grup ke-g sebelum waktu t_i

e_{gi} = nilai ekspektasi pada grup ke-g pada waktu t_i

G = banyak grup dalam satu variabel

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{\alpha, (G-1)}^2$ sehingga dapat disimpulkan bahwa minimal ada satu perbedaan kurva *survival*.

2.2 Cox Proportional Hazard Model

Cox Proportional Hazard atau dikenal dengan regresi Cox PHM adalah metode yang sering digunakan dalam analisis data *survival*. Metode ini diperkenalkan pertama kali oleh Cox (1972). Regresi Cox termasuk dalam metode analisis *survival* yang menggunakan pendekatan semi-parametrik, karena fungsi *baseline hazard* tidak mengikuti distribusi tertentu. Dalam model regresi Cox fungsi *hazard* harus proporsional atau jika digambarkan dalam kurva Kaplan-Meier tidak ada kurva yang saling berpotongan di semua nilai *survival time*.

Variabel prediktor pada model Cox dapat dinyatakan dengan suatu vektor \mathbf{x} yang memenuhi $\mathbf{x} = (x_1, x_2, \dots, x_p)^T$. Fungsi *baseline hazard* yaitu $h_0(t)$ adalah fungsi *hazard* untuk masing-masing individu yang bentuknya tidak diketahui secara spesifik. Secara umum regresi Cox dinyatakan dalam persamaan fungsi *hazard* yaitu

$$h(t, \mathbf{x}) = h_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}) \quad (2.11)$$

Atau dapat dinyatakan dalam bentuk berikut

$$h(t, \mathbf{x}) = h_0(t) \exp(\beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \dots + \beta_p \mathbf{x}_p) \quad (2.12)$$

dengan:

$h(t, \mathbf{x})$: fungsi *hazard*

\mathbf{x} : vektor variabel prediktor

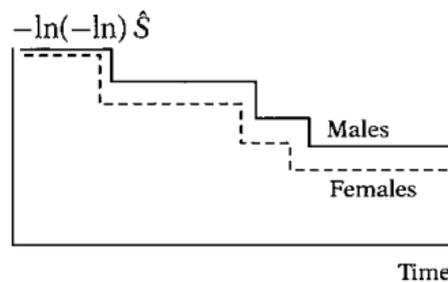
$h_0(t, \mathbf{x})$: fungsi *baseline hazard*

2.2.1 Asumsi *Proportional Hazard*

Regresi Cox PHM adalah salah satu metode untuk analisis data *survival* dengan pendekatan semi-parametrik karena pola dari *survival time* tidak diketahui atau tidak mengikuti pola tertentu. Asumsi yang harus dipenuhi regresi Cox PHM yaitu *Proportional Hazard* (PH), artinya nilai dari *hazard ratio* untuk setiap waktu bersifat tetap (*time independent*). Jika digambarkan dengan Kurva Kaplan-Meier tidak ada yang berpotongan. Untuk menguji asumsi *Proportional Hazard* ada tiga cara yang dapat dilakukan, yaitu:

a. Pendekatan Grafik

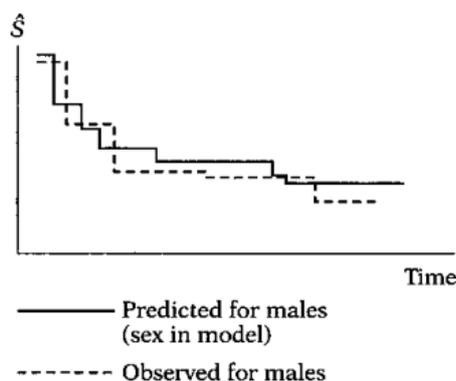
Uji asumsi PH dengan pendekatan grafik dibedakan menjadi dua, yaitu berdasarkan grafik log-log *survival* atau $-\ln(-\ln \hat{S}(t))$ dan grafik observasi-prediksi. Ilustrasi dari grafik log-log *survival* ditampilkan pada grafik berikut:



Gambar 2.4 Grafik $-\ln(-\ln \hat{S}(t))$

Pada Gambar 2.4, asumsi PH terpenuhi jika plot dari grafik $-\ln(-\ln \hat{S}(t))$ paralel dan tidak saling berpotongan. Untuk mendapatkan nilai $-\ln(-\ln \hat{S}(t))$ pada sumbu y yaitu dengan melakukan transformasi pada estimasi *survival time* ($\hat{S}(t)$). Masing-masing variabel prediktor diubah menjadi data kategori dan diuji asumsi PH secara terpisah.

Selain menggunakan grafik $-\ln(-\ln \hat{S}(t))$, pendekatan grafik juga dapat dilakukan melalui grafik observasi-prediksi dari *survival function*. Ilustrasi grafik observasi-prediksi adalah sebagai berikut:



Gambar 2.5 Grafik Observasi-Prediksi

Berdasarkan Gambar 2.5 suatu variabel prediktor memenuhi asumsi PH jika plot prediksi dan observasi sangat dekat. Masing-masing kategori pada tiap variabel prediktor diuji asumsi PH.

b. Pendekatan Uji Goodness of Fit (GOF)

Uji GOF paling banyak digunakan karena memberikan nilai *p-value* yang jelas, sehingga mampu memberikan keputusan yang lebih objektif dari uji asumsi PH. Salah satu cara uji GOF adalah dengan residual *Schoenfeld*. Langkah-langkah untuk mendapatkan residual *Schoenfeld* adalah sebagai berikut:

1. Membuat model Cox PH dan mendapatkan residual *Schoenfeld* untuk setiap variabel prediktor.
2. Membuat variabel *ranking* (v_r) berdasarkan urutan dari *survival time*. Objek yang mengalami *event* pertama kali diberi nilai 1 dan seterusnya.
3. Menguji korelasi (menggunakan korelasi rank Pearson) antara residual *Schoenfeld* yang dihasilkan pada langkah pertama dengan variabel *ranking* yang dihasilkan pada langkah kedua.

Hipotesis yang digunakan untuk menguji korelasi adalah sebagai berikut:

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

Statistika uji yang digunakan sebagai berikut:

$$t_{hit} = \frac{r_{v_r, PR_{pj}} \sqrt{n-2}}{\sqrt{1 - (r_{v_r, PR_{pj}})^2}},$$

dengan:

$$r_{v_r, PR_{pj}} = \frac{\text{cov}(v_r, PR_{pj})}{\sqrt{\text{var}(v_r) \text{var}(PR_{pj})}}$$

H_0 ditolak jika $|t_{hit}| > t_{(\alpha/2, n-2)}$, artinya asumsi PH tidak dipenuhi karena ada korelasi antara residual *Schoenfeld* yang memuat prediktor dengan ranking *survival time*.

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

$$PR_{pj} = x_{pj} - E\left(x_{pj} \mid R\left(t_{(pj)}\right)\right) \quad (2.13)$$

dimana :

$$E\left(x_{pj} \mid R\left(t_{(pj)}\right)\right) = \frac{\sum_{l \in R\left(t_{(pj)}\right)} x_{pj} \exp(\beta x_l)}{\sum_{l \in R\left(t_{(pj)}\right)} \exp(\beta x_l)} \quad (2.14)$$

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

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

$E\left(x_{pj} \mid R\left(t_{(pj)}\right)\right)$: peluang bersyarat jika $R\left(t_{(j)}\right)$ diketahui

c. Pendekatan *time dependent covariates*

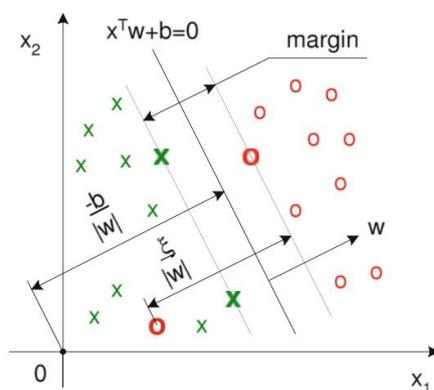
Uji asumsi PH menggunakan *time dependent covariates* adalah dengan cara menginteraksikan variabel prediktor yang akan diuji dengan fungsi waktu. Fungsi waktu yang dapat digunakan antara lain adalah fungsi linear, fungsi logaritma, dan fungsi *step*. Selanjutnya dibentuk persamaan fungsi *hazard* dan diuji variabel yang signifikan. Jika koefisien dari variabel prediktor yang diinteraksikan dengan fungsi waktu signifikan, maka asumsi PH dilanggar.

2.3 Support Vector Machine (SVM)

Support Vector Machine pertama kali diperkenalkan oleh Vapnik pada tahun 1992. Konsep dasar SVM merupakan kombinasi dari teori komputasi seperti *hyperplane*, kernel, dan pendukung lainnya. SVM termasuk dalam metode statistik yang bersifat multivariat dan non-linear serta menggunakan pendekatan *supervised learning* sehingga membutuhkan nilai respons sebagai supervisor (Gunn, 1998).

SVM memiliki dua pendekatan yaitu pendekatan klasifikasi dan pendekatan regresi. Awalnya klasifikasi dengan menggunakan SVM hanya memisahkan objek menjadi dua kelas, misalkan kelas pertama (+1) dan kelas kedua (-1). Tetapi seiring dengan berkembangnya penelitian yang dilakukan, klasifikasi dengan SVM menggunakan prinsip menemukan garis *hyperplane* optimum yang dapat memisahkan dua kelas tersebut. Banyak kemungkinan garis-garis pemisah yang terbentuk tetapi hanya ada satu garis pemisah (*hyperplane*) optimum. *Hyperplane* terbaik adalah yang mampu meminimalkan kesalahan klasifikasi dengan memaksimalkan jarak marginnya (Gunn, 1998).

Prinsip dasar SVM adalah linear *classifier* dimana terdapat suatu kasus yang dapat dipisahkan dengan baik tetapi ada juga kasus yang tidak dapat terpisah dengan baik atau terdapat kesalahan (*violation*). Secara umum, kasus yang banyak terjadi tidak dapat dipisahkan dengan baik menggunakan linear *classifier*. Ilustrasinya adalah seperti gambar berikut (Haerdle, Prasyto, dan Hafner, 2014)



Gambar 2.6 SVM dengan *Classifier* Linear

Berdasarkan Gambar 2.6 tujuan utama dari SVM adalah memaksimalkan margin sehingga jarak antar *classifier* jauh dan akan menghasilkan klasifikasi yang baik. Kondisi yang harus dipenuhi berdasarkan gambar adalah sebagai berikut:

$$\begin{cases} \mathbf{x}_i^T \mathbf{w} + b \geq 1 - \xi_i ; y_i = 1 \\ \mathbf{x}_i^T \mathbf{w} + b \leq 1 - \xi_i ; y_i = -1 \\ \xi_i \geq 0 \end{cases} \quad (2.15)$$

dimana ξ_i adalah faktor kesalahan dan jika $\xi_i > 0$ maka terjadi kesalahan klasifikasi, \mathbf{w} adalah vektor bobot dan b adalah deviasi dari titik asal.

Jika dituliskan secara lengkap fungsi tujuan dan kendalanya adalah sebagai berikut (Haerdle, Prasyto, dan Hafner, 2014).

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{i=1}^n \xi_i \quad (2.16)$$

dengan kendala :

$$\begin{aligned} \mathbf{y}_i (\mathbf{x}_i \mathbf{w} + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned}$$

Selanjutnya dibentuk fungsi *Lagrange* untuk primal problem yaitu.

$$L_p(\mathbf{w}, b, \xi) = \frac{1}{2} \|\mathbf{w}\|^2 + \gamma \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i \{ \mathbf{y}_i (\mathbf{x}_i^T \mathbf{w} + b) - 1 + \xi_i \} - \sum_{i=1}^n \mu_i \xi_i \quad (2.17)$$

Persamaan primal problem diturunkan terhadap \mathbf{w}, b, ξ_i diperoleh

$$\begin{cases} \frac{\partial L_p}{\partial \mathbf{w}} = 0 ; \mathbf{w} - \sum_{i=1}^n \alpha_i \mathbf{y}_i \mathbf{x}_i = 0 \\ \frac{\partial L_p}{\partial b} = 0 ; \sum_{i=1}^n \alpha_i \mathbf{y}_i = 0 \\ \frac{\partial L_p}{\partial \xi_i} = 0 ; \gamma - \alpha_i - \mu_i = 0 \end{cases} \quad (2.18)$$

Primal problem tidak dapat langsung diselesaikan karena ada parameter yang tidak diketahui nilainya yaitu α_i , sehingga muncul dual problem. Fungsi *Lagrange* dual problem adalah sebagai berikut:

$$L_D(\boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j \mathbf{x}_i^T \mathbf{x}_j \quad (2.19)$$

Sehingga dual problem SVM secara lengkap adalah sebagai berikut

$$\max_{\boldsymbol{\alpha}} L_D(\boldsymbol{\alpha}) \quad (2.20)$$

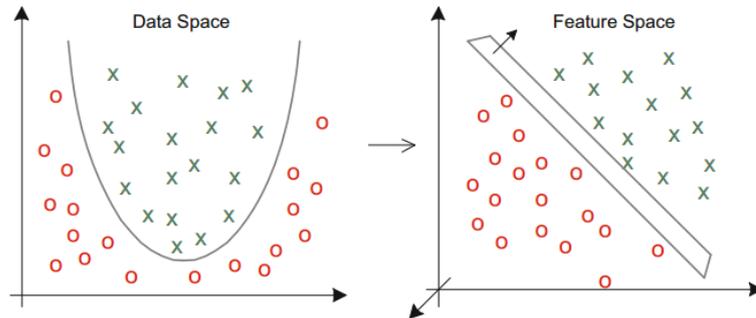
dengan kendala :

$$0 \leq \alpha_i \leq \gamma$$

$$\sum_{i=1}^n \alpha_i \mathbf{y}_i = 0.$$

Jika dilihat pada Gambar 2.6, nilai $\alpha > 0$ terjadi pada titik yang berada tepat pada garis margin, sedangkan titik-titik yang berada di luar garis margin memiliki $\alpha = 0$. Selain itu titik-titik yang berada tepat di garis margin disebut sebagai *support vector*.

Permasalahan dalam klasifikasi tidak selalu dapat diselesaikan dengan *classifier* linear, ada juga permasalahan yang diselesaikan dengan non-linear *classifier* karena data *space* nya non-linear. Ilustrasi dari non-linear SVM adalah sebagai berikut (Haerdle, Prasyto, dan Hafner, 2014).



Gambar 2.7 Non-Linear SVM

Berdasarkan Gambar 2.7 dapat dilihat bahwa terjadi transformasi dari *Data space* menjadi *Feature space* dengan menggunakan kernel. Seperti yang telah diketahui *training vectors* yaitu \mathbf{x}_i pada persamaan (2.19) merupakan perkalian skalar $\mathbf{x}_i^T \mathbf{x}_j$, maka dalam non-linear SVM perkalian skalar tersebut ditransformasikan menjadi $\varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$. Menurut Haerdle *et.al* (2014), jika fungsi kernel ada sedemikian hingga $K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j)$, maka persamaan tersebut dapat digunakan tanpa mengetahui transformasi dari φ secara eksplisit. Misalkan diketahui

$$\varphi(x_1, x_2) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2)^T, \text{ maka}$$

$$\mathbf{w}^T \varphi(x) = (w_1 \quad w_2 \quad w_3) \begin{pmatrix} x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \end{pmatrix} = w_1 x_1^2 + w_2 \sqrt{2} x_1 x_2 + w_3 x_2^2$$

Dimensi dari persamaan di atas adalah kuadrat, sehingga dengan menggunakan kernel menghindari penyelesaian secara eksplisit pada data berdimensi tinggi, sehingga

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

dimana kernel yang bersesuaian adalah

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

Beberapa jenis kernel yang sering digunakan antara lain adalah (Van Belle *et al.*, 2010a).

a. Linear Kernel

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

b. Polynomial Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\tau + \mathbf{x}_i^T \mathbf{x}_j)^a ; \tau \geq 0$$

c. RBF Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / 2\sigma^2\right)$$

Syarat dari fungsi kernel adalah definit positif, misalkan terdapat kumpulan data x_1, x_2, \dots, x_n dan sebarang bilangan real $\lambda_1, \lambda_2, \dots, \lambda_n$ fungsi kernel memenuhi kondisi berikut

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j K(\mathbf{x}_i, \mathbf{x}_j) \geq 0$$

Fungsi Kernel dalam analisis *survival* menggambarkan jarak *survival time* antar pasien. Misalkan terdapat pasien A dan pasien B yang masing-masing memiliki *survival time*. Fungsi kernel menghitung kedekatan karakteristik antara pasien A dan pasien B. Jika nilai kernel besar, maka karakteristik dari variabel

prediktor pasien A dan pasien B hampir sama. Semakin kecil nilai kernel, maka semakin berbeda karakteristiknya.

2.4 Least Squares Support Vector Machine (LS-SVM)

Least Squares Support Vector Machine (LS-SVM) pertama kali diperkenalkan oleh Suyken dan Vandewalle (1999), yang merupakan modifikasi metode SVM standard. LS-SVM mengarah pada penyelesaian linear untuk mendapatkan nilai yang optimal. Dalam formulasi LS-SVM, perhitungan komputasinya lebih sederhana jika dibandingkan dengan SVM. Fungsi tujuan dan fungsi kendala LS-SVM adalah sebagai berikut (Suyken dan Vandewalle, 1999).

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 \quad (2.21)$$

dengan kendala :

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

dan definisi fungsi kernelnya adalah $\Omega_{ij} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$.

Fungsi Lagrange dari persamaan (2.21) adalah

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 - \sum_{i=1}^n \alpha_i (y_i [\boldsymbol{\varphi}(\mathbf{x}_i)^T \mathbf{w} + b] - 1 + \xi_i) \quad (2.22)$$

dengan α_i adalah pengali Lagrange yang bisa bernilai positif atau negatif. Kondisi optimal untuk \mathbf{w} , b , α_i dapat diperoleh dengan derivative pertama dari persamaan (2.22), yaitu

$$\begin{cases} \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \boldsymbol{\varphi}(\mathbf{x}_i) \\ \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial b} = 0 \rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma \xi_i, i = 1, \dots, n \\ \frac{\partial L(\mathbf{w}, b, \alpha)}{\partial \alpha_i} = 0 \rightarrow y_i [\boldsymbol{\varphi}(\mathbf{x}_i)^T \mathbf{w} + b] = 1 - \xi_i, i = 1, \dots, n \end{cases} \quad (2.23)$$

dapat ditulis sebagai sistem linear sebagai ganti dari *Quadratic Programming* yaitu.

$$\begin{bmatrix} \mathbf{I} & 0 & 0 & -\mathbf{Z}^T \\ 0 & 0 & 0 & -\mathbf{y}^T \\ 0 & 0 & \gamma\mathbf{I} & -\mathbf{I} \\ \mathbf{Z} & \mathbf{y} & \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{b} \\ \xi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{1} \end{bmatrix} \quad (2.24)$$

dengan:

$$\mathbf{Z} = [\varphi(\mathbf{x}_1)^T y_1, \dots, \varphi(\mathbf{x}_n)^T y_n]^T, \mathbf{y} = [y_1, y_2, \dots, y_n]^T, \mathbf{1} = [1, 1, \dots, 1]^T, \xi = [\xi_1, \dots, \xi_n]^T, \\ \alpha = [\alpha_1, \dots, \alpha_n]^T, \text{ dan } \gamma \text{ adalah parameter regularisasi.}$$

2.5 Survival Support Vector Machine (SUR-SVM)

Pada prinsipnya *Survival SVM* hampir sama dengan SVM yang telah dikenal sebelumnya. Pendekatan yang digunakan dalam *Survival SVM* ada dua macam, yaitu pendekatan regresi dan pendekatan ranking (Van Belle *et al.*, 2011). Jika pada regresi Cox menggunakan fungsi *hazard*, maka pada SUR-SVM menggunakan *prognostic index* sebagai pengganti fungsi *hazard*. Prognostic dalam bidang kedokteran diartikan sebagai perkiraan hasil penyembuhan pasien berdasarkan pada *treatment* atau perawatan medis yang telah dilakukan atau dapat juga dilihat berdasarkan faktor intrinsik dari penyakit yang dialami, misalkan ukuran sel kanker dan pengaruh pemberian hormon terhadap jumlah sel kanker (Galea, Blamey, Elston, dan Ellis, 1992).

Menurut Van Belle *et al.* (2010a) fungsi *prognostic* atau disebut juga fungsi utilitas dari model *survival SVM* yang sering digunakan dalam bidang medis yaitu $u: \mathbb{R}^d \rightarrow \mathbb{R}$ yang didefinisikan sebagai

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

dengan \mathbf{w} adalah suatu vektor parameter dan $\boldsymbol{\varphi}(x)$ merupakan transformasi (*feature map*) dari kovariat x dimana berdasarkan *Mercer's condition* bentuk dari *feature map* boleh tidak terdefinisi saat menggunakan kernel yang memenuhi sifat definit positif. Fungsi tujuan dan fungsi kendala dari SURSVM adalah sebagai berikut (Van Belle *et al.*, 2010a)

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_i \sum_{j, i < j} v_{ij} \xi_{ij} ; \gamma \geq 0 \quad (2.25)$$

dengan kendala :

$$\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) \geq 1 - \xi_{ij} ; \forall i < j$$

$$\xi_{ij} \geq 0 ; \forall i < j$$

dimana v_{ij} merupakan indikator pembandingan antara pasien ke- i ($\mathbf{x}_i, t_i, \delta_i$) dan pasien ke- j ($\mathbf{x}_j, t_j, \delta_j$) yang memenuhi :

$$v_{ij} = \begin{cases} 1; (t_i < t_j, \delta_i = 1) \\ 0; (t_i < t_j, \delta_i = 0) \end{cases} \quad (2.26)$$

dan *slack* variabel (ξ_{ij}) merupakan nilai pelanggaran jika terjadi kesalahan ranking *failure times*. Untuk indikator pembandingan, misalkan terdapat pasien i dan pasien j yang masing-masing memiliki *survival time* berturut-turut t_i dan t_j serta status tersensor berturut-turut δ_i dan δ_j . v_{ij} digunakan untuk mengetahui apakah dua pasien dapat dibandingkan atau tidak. Dua pasien dapat dibandingkan jika kedua pasien mengalami event ($\delta = 1$) atau satu pasien dengan *survival time* lebih kecil mengalami event dan satu pasien lainnya tersensor kanan ($v_{ij} = 1$). Jika kedua pasien tersensor kanan maka $t_i = t_j$ akibatnya $v_{ij} = 0$.

Perbedaan antara *Survival SVM* dengan model Cox adalah pada *Survival SVM* tidak mengasumsikan linearitas pada variabel kovariat seperti pada model Cox. Selain itu, *Survival SVM* dapat digunakan untuk data berdimensi tinggi (*high-dimensional data*) dan menjamin bahwa solusi yang diperoleh memenuhi global optimum karena bentuk fungsinya konveks (Van Belle *et al.*, 2010b).

2.6 *Survival Least Squares-Support Vector Machine (SURLS-SVM)*

Van Belle *et al.* (2010a) memperkenalkan metode baru dalam analisis *survival* yaitu *Survival LS-SVM*. Metode ini mengadopsi metode yang telah diperkenalkan Suyken dan Vandewalle (1999) yaitu LS-SVM. Dalam metode *Survival LS-SVM* menggunakan pendekatan ranking dari *survival time* pasien. Keunggulan dari metode ini jika dibandingkan dengan SVM biasa adalah pada LS-SVM menggunakan pendekatan linear pada penyelesaian sistem KKT sehingga mempermudah dalam komputasi. Hal ini sangat membantu jika diterapkan pada data berdimensi tinggi. Fungsi tujuan dan kendala untuk *Survival LS-SVM* adalah sebagai berikut

$$\min_{\mathbf{w}, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 \quad (2.27)$$

dengan kendala

$$\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) = 1 + \xi_{ij} ; \forall i, j = 1, 2, \dots, n$$

dimana $\gamma > 0$ adalah parameter reguralisasi. Perbedaan antara SURSVM dengan SURLS-SVM adalah pada model kedua ini fungsi kendala berupa persamaan.

Persamaan Lagrange untuk model SURLS-SVM adalah sebagai berikut:

$$L(\mathbf{w}, \xi, \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^n \sum_{j=1}^n v_{ij} \xi_{ij}^2 - \sum_{i=1}^n \sum_{j=1}^n \boldsymbol{\alpha}_{ij} (\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - 1 - \xi_{ij}) \quad (2.28)$$

Nilai optimal dari parameter yang ada pada persamaan Lagrange akan diselesaikan dengan derivatif pertama yaitu

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} = \sum_{i,j=1}^n \boldsymbol{\alpha}_{ij} (\boldsymbol{\varphi}(\mathbf{x}_j) - \boldsymbol{\varphi}(\mathbf{x}_i)) \\ \frac{\partial L}{\partial \xi_{i,j}} = 0 \rightarrow \boldsymbol{\alpha}_{ij} = \gamma v_{ij} \xi_{ij} \\ \frac{\partial L}{\partial \boldsymbol{\alpha}_{ij}} = 0 \rightarrow \mathbf{w}^T (\boldsymbol{\varphi}(\mathbf{x}_j) - \boldsymbol{\varphi}(\mathbf{x}_i)) - 1 - \xi_{ij} = 0 \end{cases} \quad (2.29)$$

Selanjutnya dengan mengeliminasi \mathbf{w} dan ξ_{ij} akan diperoleh suatu persamaan linear berikut

$$[\gamma \mathbf{DKD}^T + \mathbf{I}_{n^2}] \boldsymbol{\alpha} = \gamma \mathbf{1}_{n^2/2} \quad (2.30)$$

dimana \mathbf{D} adalah matriks yang memiliki elemen $\{-1,0,1\}$, yang didefinisikan sedemikian hingga memenuhi persamaan berikut:

$$\mathbf{DX} = \begin{pmatrix} x_1 - x_2 \\ \vdots \\ x_1 - x_n \\ \vdots \\ x_{n-1} - x_n \end{pmatrix}$$

dengan $\mathbf{X} = (x_1, x_2, \dots, x_n)^T$ dan \mathbf{D} memiliki ukuran $n_c \times n$ dengan n_c adalah banyaknya objek yang dapat dibandingkan dan n adalah banyaknya objek yang diamati. γ adalah parameter regularisasi dan \mathbf{I} adalah matriks identitas, dan

$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}_j)$ merupakan elemen dari kernel matriks. Model yang diperoleh untuk nilai x yang baru (x^*) adalah sebagai berikut:

$$\begin{aligned}
\hat{\mathbf{u}}^* &= \hat{\mathbf{w}}^T \boldsymbol{\varphi}(\mathbf{x}^*) \\
&= \sum_{ij} \hat{\boldsymbol{\alpha}}_{ij} (\boldsymbol{\varphi}(\mathbf{x}_j) - \boldsymbol{\varphi}(\mathbf{x}_i))^T \boldsymbol{\varphi}(\mathbf{x}^*) \\
&= \hat{\boldsymbol{\alpha}}^T \mathbf{D} \mathbf{K}_n(\mathbf{x}^*)
\end{aligned} \tag{2.31}$$

dengan $\mathbf{K}_n(\mathbf{x}^*) = [\mathbf{K}(\mathbf{x}_1, \mathbf{x}^*) \dots \mathbf{K}(\mathbf{x}_n, \mathbf{x}^*)]^T$, $\hat{\mathbf{u}}^*$ adalah *prognostic index*, dan $\hat{\boldsymbol{\alpha}}$ adalah solusi dari persamaan (2.30). Dengan menggunakan *prognostic index* dapat diketahui performansi dari model SURLS-SVM. Model SURLS-SVM pada persamaan (2.27) merupakan model yang menggunakan pendekatan ranking antara *survival time* dengan *prognostic index*, hal ini bertujuan untuk mengetahui pasien mana yang berada pada risiko tinggi dan pasien mana yang berada pada risiko rendah.

2.7 Kriteria Kebaikan Model

Salah satu kriteria kebaikan model adalah c-indeks (*concordance index*) yang pertama kali diperkenalkan oleh Frank (1984), yang menyatakan bahwa c-index merupakan proporsi dari semua pasangan *survival time* yang dapat digunakan untuk menentukan urutan *survival time* sehingga akan menghasilkan prediksi yang tepat. c-indeks ini mengukur keterurutan antara fungsi *prognostic* dan *survival time* observasi baik untuk data tersensor maupun data yang tidak tersensor atau mengukur missranking. Semakin besar nilai c-index akan memberikan performansi yang semakin baik (Mahjub *et al.*, 2016). Persamaan empiris yang menyatakan besarnya c-index pada dataset $\mathcal{D} = \{(x_i, t_i, \delta_i)\}_{i=1}^n$ adalah sebagai berikut (Van Belle *et al.*, 2008).

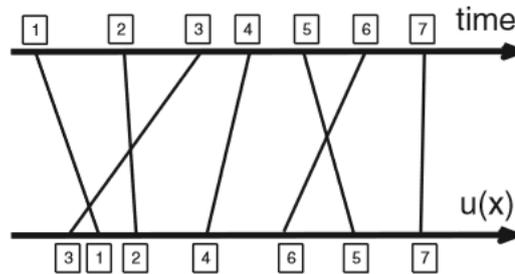
$$c_{ij}(\mathbf{u}) = \frac{\sum_{i=1}^n \sum_{i < j} v_{ij} I((u(\mathbf{x}_j) - u(\mathbf{x}_i))(t_j - t_i) > 0)}{\sum_{i=1}^n \sum_{i < j} v_{ij}} \tag{2.32}$$

dimana I adalah fungsi indikator yang didefinisikan sebagai berikut

$$I = \begin{cases} 1; & (u(\mathbf{x}_j) - u(\mathbf{x}_i))(t_j - t_i) > 0 \\ 0; & \text{lainnya} \end{cases}$$

dengan v_{ij} adalah indikator pembandingan seperti pada persamaan (2.26), $u(\mathbf{x})$ adalah *prognostic index*, dan t adalah *survival time*. Keterurutan yang baik diperoleh dengan menyelesaikan kendala, dan pelanggaran yang disebabkan oleh misranking

antara pasangan observasi (Van Belle *et al.*, 2010a). Pada model SUR-SVM dan SURLS-SVM digunakan pendekatan ranking sehingga cara menghitung c-index adalah dengan menentukan proporsi antara pasangan objek yang tidak missranking dengan semua pasangan objek yang mungkin. Penghitungan c-index dapat diilustrasikan dalam gambar berikut (Van Belle *et al.*, 2011).



Gambar 2.8 Ilustrasi penghitungan c-index

Berdasarkan Gambar 2.8 dapat dibuat tabel pengamatan yang missranking dan yang tidak missranking sebagai berikut

concordance	rank						ALL
	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
ALL	5	4	4	3	1	1	c-index 18/21

Berdasarkan tabel di atas nilai 1 adalah indikator untuk pasangan objek yang tidak missranking dan nilai 0 adalah indikator untuk pasangan objek yang missranking (salah urutan). Misal pasangan objek 1 dan 2 berdasarkan *survival time* $t_1 < t_2$, dan berdasarkan *prognostic index* ($u(x)$) yaitu $u(x_1) < u(x_2)$ sehingga antara *survival time* dan *prognostic index* sesuai (bernilai 1). Sedangkan pada pasangan objek 1 dan 3 berdasarkan *survival time* $t_1 < t_3$ tetapi berdasarkan *prognostic index* $u(x_1) > u(x_3)$ sehingga terjadi missranking pada *prognostic index* (bernilai 0). Banyaknya pasangan yang mungkin ditentukan dengan kombinasi. Pada contoh di atas ada 7 objek sehingga dengan menggunakan kombinasi diperoleh 21 kemungkinan pasangan. Dari ringkasan tabel diketahui bahwa ada 18 objek yang tidak miss-ranking sehingga nilai c-index = $\frac{18}{21}$.

Selain menggunakan c-index, kebaikan model dapat juga dilihat berdasarkan log-rank. Menurut Mahjub *et al.* (2016) nilai log-rank digunakan untuk

menentukan apakah model mampu memisahkan pasien yang masuk kriteria *high risk* dan pasien yang masuk *low risk*. Model mampu memisahkan jika hasil log-rank test menunjukkan bahwa dua kriteria itu signifikan berbeda. Dasar yang digunakan untuk memisahkan dua kategori dapat menggunakan mean atau median dari *prognostic index*, selanjutnya diuji dengan menggunakan uji log-rank (Mahjub *et al.*, 2016).

Kriteria lain yang dapat digunakan adalah *hazard ratio*. Pada model SURSVM dan SURLSSVM untuk mendapatkan *hazard ratio* tidak bisa secara langsung seperti pada model Cox. Menurut Van Belle *et al.* (2011) untuk mendapatkan *hazard ratio* diperoleh dengan mengestimasi melalui model Cox. Setelah mendapatkan $u(\mathbf{x})$ yang merupakan *prognostic index*, selanjutnya masing-masing $u(\mathbf{x})$ dinormalisasi dan memiliki nilai antara 0 sampai dengan 1. Selanjutnya dibentuk menjadi model Cox sehingga *hazard ratio* dapat ditentukan. Semakin besar nilai *hazard ratio* semakin baik.

2.8 Pembangkitan Data *Survival Time*

Studi simulasi merupakan salah satu cara untuk merepresentasikan kinerja, sifat, dan kecukupan model dalam situasi tertentu. Model yang sering digunakan untuk pendekatan dalam simulasi data *survival* adalah model Cox, sedangkan salah satu distribusi yang sering digunakan adalah Weibull. Diantara semua distribusi yang dapat dipakai untuk data *survival*, hanya eksponensial, Weibull, dan Gompertz yang menjamin sifat PH dapat terpenuhi (Bender *et al.*, 2005). Parameter Weibull dapat dipilih sedemikian rupa sehingga asumsi PH terpenuhi dan nilai dari *hazard ratio* dapat dihitung berdasarkan parameter Weibull. Fungsi kepadatan peluang dari distribusi Weibull adalah (Bender *et al.*, 2005)

$$f(t) = \lambda \nu t^{\nu-1} \exp(-\lambda t^\nu) \quad (2.33)$$

dimana $\lambda > 0$ adalah parameter skala, $\nu > 0$ adalah parameter bentuk, $t > 0$ adalah *survival time*. Data yang dibangkitkan dengan distribusi Weibull bernilai positif atau berada pada range $[0, \infty)$. Berdasarkan fungsi kepadatan peluang pada persamaan (2.33) dapat diturunkan persamaan untuk fungsi *survival* dan fungsi

hazard. Fungsi *survival* untuk distribusi Weibull adalah sebagai berikut (Bender *et al.*, 2005)

$$S_0(t) = \exp(-\lambda t^\nu) \quad (2.34)$$

sedangkan fungsi *baseline hazard* untuk distribusi Weibull adalah sebagai berikut

$$h_0(t) = \lambda \nu t^{\nu-1} \quad (2.35)$$

Fungsi *survival* Cox PH didefinisikan sebagai berikut

$$S(t, \mathbf{x}) = \exp[-H_0(t) \exp(\boldsymbol{\beta}^T \mathbf{X})] \quad (2.36)$$

dengan $H_0(t)$ adalah fungsi kumulatif *hazard* yang merupakan integral dari fungsi *baseline hazard* $h_0(t)$ sehingga didapatkan persamaan untuk $H_0(t) = \lambda t^\nu$ dan $H_0^{-1}(t) = (\lambda^{-1}t)^{1/\nu}$. Misalkan $S(t, \mathbf{x}) = U$, maka persamaan (2.36) dapat dituliskan menjadi

$$U = \exp[-H_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x})] \quad (2.37)$$

Untuk mendapatkan persamaan *survival time* dilakukan dengan menginverskan fungsi $H_0(t)$, dan sebelumnya diubah menjadi persamaan berikut

$$\log U = -H_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}) \quad (2.38)$$

maka persamaan untuk *survival time* dengan pendekatan Cox Weibull adalah

$$T = -H_0^{-1} \left[\frac{\log U}{\exp(\boldsymbol{\beta}^T \mathbf{x})} \right] \quad (2.39)$$

Misalkan $t = \frac{\log U}{\exp(\boldsymbol{\beta}^T \mathbf{x})}$, maka dengan substitusi invers dari $H_0(t)$ ke persamaan

(2.39) diperoleh

$$T = \left(-\frac{\log(U)}{\lambda \exp(\boldsymbol{\beta}^T \mathbf{x})} \right)^{1/\nu} \quad (2.40)$$

Fungsi *hazard* dapat diperoleh dengan menggunakan persamaan (2.7) yaitu

$$\begin{aligned} h(t, \mathbf{x}) &= -\frac{d \ln (S(t, \mathbf{x}))}{dt} \\ &= -\frac{d}{dt} \ln [\exp(-\lambda t^\nu \exp(\boldsymbol{\beta}^T \mathbf{x}))] \\ &= \frac{d}{dt} [\lambda t^\nu \exp(\boldsymbol{\beta}^T \mathbf{x})] \end{aligned}$$

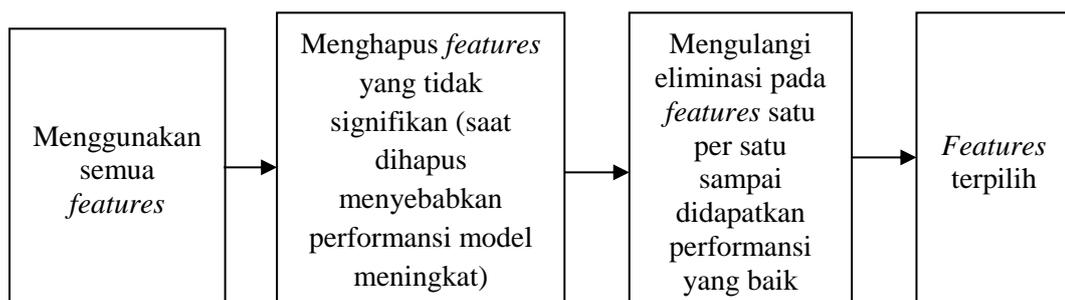
$$h(t | \mathbf{x}) = \lambda \exp(\boldsymbol{\beta}^T \mathbf{x}) \nu t^{\nu-1} \quad (2.41)$$

dimana $U \sim U(0,1)$. Berdasarkan persamaan-persamaan yang sudah diketahui, maka dengan mengatur nilai parameter Weibull dapat dibangkitkan *survival time* yang berdistribusi Weibull.

2.9 Metode *Feature Selection (Backward Elimination)*

Feature merupakan variabel terukur yang dapat menentukan performansi dari suatu pengamatan. Tujuan dari *feature selection* adalah memilih *feature* yang mampu merepresentasikan kumpulan *feature* sehingga mengurangi kemungkinan adanya *feature* yang tidak relevan di dalam model (Chandrashekar dan Sahin, 2014).

Feature selection yang banyak digunakan antara lain adalah *wrapper method* dan *filter method*. Salah satu jenis *wrapper method* adalah *backward elimination*. Dalam *backward elimination* yang digunakan dalam *machine learning* langkah awalnya adalah mencari nilai performansi dari suatu model saat semua *features* masih ada dalam model, selanjutnya setiap *features* yang membuat nilai performansi naik saat *feature* itu dihapus maka itulah yang keluar dari model (Chandrashekar dan Sahin, 2014). *Backward elimination* dipilih karena dapat mendeteksi variabel yang signifikan jika bersama-sama ada dalam model tetapi tidak signifikan ketika secara individu ada dalam model (*suppresor variable*) artinya variabel-variabel tersebut akan menyebabkan c-index naik ketika secara bersama-sama ada dalam model, dan *forward elimination* tidak dapat mendeteksi ini karena pada metode *forward elimination* cara memasukkan variabel dalam model adalah dilihat variabel yang memiliki kenaikan c-index besar saat variabel itu secara individu ada dalam model. Tahapan dari *backward elimination* adalah sebagai berikut:



Gambar 2.9 Tahapan *Backward Elimination*

2.10 Kanker Serviks

Kanker merupakan suatu penyakit yang disebabkan oleh kelainan siklus sel, dimana sel tumbuh tidak terkendali dan menyerang jaringan biologis. Kanker juga dapat menyebar ke seluruh jaringan tubuh melalui kelenjar getah bening maupun pembuluh darah. Sel kanker ini mengalami perubahan dalam unit penyusunnya (DNA) sehingga terjadi mutasi. Mutasi ini menyebabkan sel terus membelah diri dan tidak lagi dapat merespon sinyal yang dikirimkan badan untuk menghentikan pembelahan, dan dari proses ini terbentuk massa sel dari satu jenis yang disebut tumor. Menurut Samadi (2010) kanker serviks adalah kanker yang tumbuh dan berkembang pada serviks atau mulut rahim, khususnya berasal dari lapisan epitel atau lapisan terluar permukaan serviks. Tingkat keparahan dari kanker serviks dapat dilihat salah satunya berdasarkan stadium atau *stage* dari kanker serviks. Menurut FIGO (*The International Federation of Gynecology and Obstetrics*) Staging untuk kanker serviks adalah sebagai berikut:

Tabel 2.1 Staging Kanker Serviks

Stage	Deskripsi
0	Karsinoma setempat : terlihat perubahan seperti kanker, tetapi tanpa hasil temuan bahwa kanker telah menyebar ke jaringan sekelilingnya (stroma).
I	Kanker terbatas pada serviks. Ia Kanker yang menyebar dengan penyebaran ke dalam jaringan penopang tidak lebih dari 5mm dalamnya dan 7mm lebarnya. Ia1 Penyebaran ke dalam jaringan penopang dalamnya 3mm atau kurang. Ia2 Ukuran penyebaran ke dalam jaringan penopang lebih dari 3mm dan kurang dari 5mm. Ib Lesi di serviks terlihat dalam pengamatan atau menunjukkan penyebaran ke dalam jaringan penopang lebih luas daripada lesi Ia. Ib1 Tumor kurang dari 4cm. Ib2 Tumor lebih dari 4cm.
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 pinggul 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 (misal : paru-paru dan hati).

Sumber : WHO (2005)

Pengertian dari leher rahim (serviks atau *cervix*) adalah bagian rahim yang menonjol ke dalam lengkungan vagina yang terdiri dari pars *vaginalis* atau partio dan *pars supra vaginalis uteri* atau bagian kanalis yang berada di atas saluran vagina yang berada pada leher rahim, panjangnya 2,5 cm yang dilapisi kelenjar-kelenjar bersilia yang berfungsi sebagai *reseptakulum seminis* dengan pintu saluran leher rahim sebelah dalam dan pintu saluran leher rahim di vagina (Prawirohardjo, 2001).

Berdasarkan biopsi yang dilakukan diketahui bahwa proses perubahan sel kanker dari jinak menjadi ganas memerlukan waktu yang cukup lama yaitu 10-15 tahun. Penyebaran kanker serviks melalui getah bening secara umum menuju ke tiga arah, yaitu ke arah forniks dan dinding vagina, ke arah korpus uteri, ke arah parametrium dan dalam tingkatan yang lanjut menginfiltrasi *septum rectovaginal* dan kandung kemih (Kustiyati dan Winarni, 2011).

Dalam beberapa penelitian yang pernah dilakukan mengenai kanker serviks, hal yang menjadi fokus penelitian adalah model dari fungsi *hazard*, seperti pada penelitian yang dilakukan oleh Purnami *et.al.* (2015) dan Respita, R.D. (2017). Padahal selain fungsi *hazard* ada hal lain yang dapat digunakan sebagai fokus penelitian yaitu *prognostic index* dari pasien kanker serviks. Menurut Galea *et.al.* (1992) *prognostic index* merupakan salah satu ukuran yang dapat digunakan untuk melihat keberhasilan dari suatu terapi yang telah dijalani pasien kanker. Pada penelitian yang dilakukan oleh Endo *et.al.* (2015), *prognostic index* kanker serviks ditentukan oleh beberapa faktor klinis antara lain usia pasien, stadium, sejarah tumor, diameter tumor, jenis kemoterapi yang dilakukan, jumlah hemoglobin sebelum dilakukan treatment, dan ukuran metastasis.

Penelitian yang lain mengenai kanker serviks menyebutkan bahwa wanita yang berusia 30-54 tahun memiliki resiko terkena kanker serviks lebih tinggi dibandingkan dengan wanita usia 55-64 tahun dan 65 tahun ke atas (Ioka *et al.*, 2005). Lee dan Kim (2013) menyimpulkan bahwa dari 38 kasus kambuh pada kanker serviks, pasien yang menjalani pengobatan operasi dengan *adjuvant* CCRT (*Concurrent Chemo-Radiation Therapy*) cenderung lebih beresiko dibandingkan dengan hanya menjalani operasi saja dan operasi dengan *adjuvant* RT (*Radiotherapy*).

Menurut Dzion *et al.* (2011) indikator perkiraan perkembangan penyakit adalah hasil temuan dari evaluasi tumor kanker yang dapat memberikan sedikit informasi seberapa lama penderita dapat bertahan dan seberapa besar peluang seorang pasien sembuh. Indikator perkiraan paling kuat adalah tabel stadium yang menggambarkan sejauh mana penyebaran dari sel kanker. Secara medis semakin tinggi stadium, maka kemungkinan pasien untuk sembuh semakin menurun.

Beberapa faktor yang berkaitan dengan kanker serviks menurut Diananda (2007) adalah sebagai berikut :

a. usia

usia yang beresiko tinggi menderita kanker serviks adalah lebih dari 35 tahun. Semakin tua usia seseorang, semakin tinggi resiko terkena kanker serviks karena sistem imun tubuh yang menurun dan bertambah lamanya paparan terhadap karsinogen,

b. usia pertama kali menikah

menikah terlalu dini yaitu di bawah 20 tahun beresiko terkena kanker serviks 10-12 kali lebih besar daripada wanita yang menikah di atas 20 tahun. Saat wanita berusia di bawah 20 tahun dianggap belum matang untuk melakukan hubungan seksual karena kematangan tidak hanya dilihat berdasarkan sudah menstruasi atau belum tetapi juga dilihat berdasarkan sel-sel mukosa yang terdapat di selaput kulit bagian dalam rongga tubuh. Umumnya sel-sel mukosa benar-benar matang setelah seorang wanita berusia di atas 20 tahun. Jadi seorang wanita yang melakukan hubungan seks di bawah 20 tahun masih rentan terhadap rangsangan sehingga tidak siap menerima rangsangan dari luar termasuk zat-zat kimia yang dibawa oleh sperma. Karena masih rentan inilah yang menyebabkan sel mukosa berubah menjadi sel kanker,

c. terinfeksi *Human Papilloma Virus* (HPV)

infeksi virus HPV dapat berasal dari seringnya berganti-ganti pasangan. Virus ini mengubah sel-sel mukosa hingga membelah lebih banyak dan tidak terkendali sehingga menjadi kanker,

d. penggunaan antiseptik

kebiasaan menggunakan zat-zat antiseptik untuk membersihkan vagina dapat menyebabkan iritasi di serviks,

e. merokok

nikotin yang ada dalam rokok merangsang sel-sel tubuh bereaksi, baik pada mukosa tenggorokan, paru-paru, maupun serviks. Wanita perokok beresiko 4-13 kali terkena kanker serviks daripada wanita bukan perokok,

f. paritas (jumlah kelahiran)

wanita yang sering melahirkan akan memiliki resiko terkena kanker serviks lebih besar karena seringnya terjadi luka pada organ reproduksi saat persalinan, apalagi jika jarak persalinannya pendek. Luka yang terjadi akan memudahkan terkena virus HPV,

g. penggunaan kontrasepsi oral

penggunaan kontrasepsi oral yang dipakai dalam jangka waktu lebih dari 4 tahun akan meningkatkan resiko kanker serviks sebesar 1,5 – 2,5 kali. Hal ini karena jaringan leher rahim adalah salah satu sasaran hormon steroid yang terkandung dalam kontrasepsi oral.

Untuk melakukan pencegahan penyebaran sel kanker serviks, dapat dimulai dari diri sendiri yaitu dengan mengenali gejala-gejala yang nampak, antara lain (Kustiyati dan Winarni, 2011):

- a. masa tanpa gejala, pada masa ini penderita tidak mengeluh dan tidak merasakan suatu gejala meskipun sebenarnya pasien sudah mengidap penyakit kanker serviks,
- b. keputihan merupakan gejala yang sering ditemukan. Keputihan wajar terjadi pada wanita jika tidak berbau busuk, namun pada penderita kanker serviks, keputihan yang dialami makin lama makin banyak dan berbau busuk karena infeksi dan nekrosis jaringan,
- c. pendarahan yang terjadi lebih sering dan sulit berhenti akibat terbukanya pembuluh darah,
- d. rasa nyeri yang terjadi karena infiltrasi sel tumor ke serabut saraf,
- e. anemia yang terjadi karena pendarahan yang lebih sering terjadi,
- f. gejala yang dapat timbul karena metastatis jauh, mislanya cepat lelah dan penurunan berat badan secara signifikan.

(Halaman ini sengaja dikosongkan)

BAB 3

METODE PENELITIAN

Metode penelitian membahas mengenai metode untuk mengkaji SURLS-SVM dan aplikasinya pada data simulasi dan data pasien kanker serviks. Masing-masing akan dijelaskan dalam subbab berikut.

3.1 Kajian Teoritis Model SURLS-SVM

Tujuan pertama pada penelitian ini adalah melakukan kajian model SURLS-SVM untuk mendapatkan *prognostic index* seperti pada persamaan (2.31). Langkah penelitian yang dilakukan untuk mencapai tujuan pertama adalah sebagai berikut.

1. Diberikan dataset $\mathcal{D} = \{(\mathbf{x}_i, t_i, \delta_i)\}_{i=1}^n$ dengan \mathbf{x}_i adalah variabel prediktor untuk objek ke- i , t_i adalah *survival time* objek ke- i , dan δ_i adalah status tersensor objek ke- i ($\delta_i = 1$ jika terjadi *event* dan $\delta_i = 0$ jika tersensor atau tidak terjadi *event*).
2. Memasukkan persamaan kernel, dimana yang digunakan dalam penelitian ini adalah *Radial Basis Function* yaitu

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{2\sigma^2}\right)$$

dengan :

\mathbf{x}_i : variabel prediktor pasien ke- i

σ^2 : parameter regulasi

\mathbf{x}_j : variabel prediktor pasien ke- j

3. Menyelesaikan optimasi SURLS-SVM yaitu sebagai berikut.

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

dengan kendala

$$\mathbf{w}^T \phi(\mathbf{x}_j) - \mathbf{w}^T \phi(\mathbf{x}_i) = 1 + \xi_{ij} ; \forall i, j = 1, \dots, n$$

Dalam penelitian ini *feature map* yang digunakan untuk mentransformasi dimensi data adalah *kernel trick*.

4. Menyusun persamaan Lagrangian berdasarkan fungsi yang akan dioptimalkan. Misal diberikan $g(x)$ adalah fungsi tujuan, $h(x)$ adalah fungsi kendala, dan λ adalah pengali Lagrange, maka persamaan Lagrangian dapat dituliskan sebagai $L(w, \xi, \alpha) = g(x) - \lambda h(x)$.
5. Melakukan derivatif parsial dari persamaan Lagrange terhadap parameter \mathbf{w} , ξ , α , dan disamadengkan nol untuk mendapatkan nilai yang optimal untuk masing-masing parameter.
6. Berdasarkan hasil pada langkah 5, selanjutnya dilakukan substitusi dan eliminasi parameter \mathbf{w} dan ξ_{ij} .
7. Mendapatkan taksiran untuk parameter α yaitu $\hat{\alpha}$ berdasarkan hasil dari langkah 6.
8. Menyusun persamaan prediksi *prognostic index* berdasarkan persamaan berikut

$$\hat{\mathbf{u}}(\mathbf{x}^*) = \hat{\mathbf{w}}^T \mathbf{K}(\mathbf{x}^*)$$

dengan $\hat{\mathbf{u}}(\mathbf{x}^*)$ adalah *prognostic index* dan \mathbf{x}^* adalah titik pengamatan yang baru

3.2 Aplikasi Model SURLS-SVM

Setelah terbentuk *prognostic index* yang merupakan pengganti fungsi *hazard*, selanjutnya model SURLS-SVM diaplikasikan pada data simulasi dan data riil (data pasien kanker serviks di RS dr. Soetomo Surabaya).

3.2.1 Data Simulasi

Data simulasi digunakan untuk menunjukkan performansi model SURLS-SVM pada berbagai ukuran data survival. Data simulasi yang digunakan dibedakan menjadi dua jenis data, yaitu data parametrik dan non-parametrik. Ukuran data yang digunakan untuk masing-masing jenis data adalah 100, 1.000, dan 5.000. Untuk status tersensor dari variabel respon digunakan beberapa nilai yaitu 0.1, 0.2, ..., 0.9. Variabel yang dibangkitkan adalah variabel prediktor dan variabel respon. Skenario yang digunakan dalam membangkitkan data yaitu sebagai berikut.

A. *Survival time* (T) yang merupakan variabel respon dibangkitkan mengikuti distribusi Weibull berbasis model Cox dengan persamaan untuk *survival time* sebagai berikut:

$$T = \left(-\frac{\log(U)}{\lambda \exp(\boldsymbol{\beta}' \mathbf{x})} \right)^{1/\nu},$$

dengan nilai $\lambda = 2$, $\nu = 2$, dan $U \sim U(0,1)$. Distribusi untuk membangkitkan prediktor serta nilai koefisien β dirangkum pada Tabel 3.1.

Tabel 3.1 Distribusi Variabel Prediktor

Jumlah Prediktor	Distribusi	Nilai Parameter		
$p = 6$	$X_1 \sim BIN(1, 0.5)$	$X_5 \sim N(40,3)$	$\beta_1 = 0.01$	$\beta_{41} = 0.015$
	$X_2 \sim BIN(1, 0.3)$	$X_6 \sim N(25,2)$	$\beta_2 = -0.015$	$\beta_{42} = 0.01$
	$X_3 \sim Mult(4, 0.2, 0.3, 0.4, 0.1)$		$\beta_{31} = -0.07$	$\beta_{43} = -0.03$
	$X_4 \sim Mult(4, 0.5, 0.1, 0.2, 0.2)$		$\beta_{32} = 0.04$	$\beta_5 = -0.15$
			$\beta_{33} = -0.08$	$\beta_6 = -0.08$
$p = 8$	$X_1 \sim BIN(1, 0.5)$	$X_7 \sim N(20,3)$	$\beta_1 = 0.01$	$\beta_{42} = 0.01$
	$X_2 \sim BIN(1, 0.3)$	$X_8 \sim N(35,2)$	$\beta_2 = -0.015$	$\beta_{43} = -0.03$
	$X_3 \sim Mult(4, 0.2, 0.3, 0.4, 0.1)$		$\beta_{31} = -0.07$	$\beta_5 = -0.15$
	$X_4 \sim Mult(4, 0.5, 0.1, 0.2, 0.2)$		$\beta_{32} = 0.04$	$\beta_6 = -0.08$
	$X_5 \sim N(40,3)$		$\beta_{33} = -0.08$	$\beta_7 = 0$
	$X_6 \sim N(25,2)$		$\beta_{41} = 0.015$	$\beta_8 = 0$
$p = 15$	$X_1 \sim BIN(1, 0.5)$	$X_{11} \sim N(40,3)$	$\beta_1 = 0.01$	$\beta_{81} = -0.028$
	$X_2 \sim BIN(1, 0.3)$	$X_{12} \sim N(25,2)$	$\beta_2 = -0.015$	$\beta_{82} = 0.05$
	$X_3 \sim BIN(1, 0.7)$	$X_{13} \sim N(20,3)$	$\beta_3 = 0.015$	$\beta_{91} = 0.03$
	$X_4 \sim BIN(3, 0.4)$	$X_{14} \sim N(35,2)$	$\beta_4 = -0.021$	$\beta_{92} = -0.08$
	$X_5 \sim BIN(1, 0.2)$	$X_{15} \sim N(17,2)$	$\beta_5 = -0.05$	$\beta_{101} = 0.04$
	$X_6 \sim Mult(4, 0.2, 0.3, 0.4, 0.1)$		$\beta_{61} = -0.07$	$\beta_{102} = -0.018$
	$X_7 \sim Mult(4, 0.5, 0.1, 0.2, 0.2)$		$\beta_{62} = 0.04$	$\beta_{11} = -0.15$
	$X_8 \sim Mult(3, 0.3, 0.1, 0.6)$		$\beta_{63} = -0.08$	$\beta_{12} = -0.08$
	$X_9 \sim Mult(3, 0.2, 0.4, 0.4)$		$\beta_{71} = 0.015$	$\beta_{13} = -0.01$
	$X_{10} \sim Mult(3, 0.7, 0.2, 0.1)$		$\beta_{72} = 0.01$	$\beta_{14} = -0.02$
		$\beta_{73} = -0.03$	$\beta_{15} = -0.075$	
$p = 17$	$X_1 \sim BIN(1, 0.5)$	$X_{13} \sim N(20,3)$	$\beta_1 = 0.01$	$\beta_{82} = 0.05$
	$X_2 \sim BIN(1, 0.3)$	$X_{14} \sim N(35,2)$	$\beta_2 = -0.015$	$\beta_{91} = 0.03$
	$X_3 \sim BIN(1, 0.7)$	$X_{15} \sim N(17,2)$	$\beta_3 = 0.015$	$\beta_{92} = -0.08$
	$X_4 \sim BIN(1, 0.4)$	$X_{16} \sim N(50, 1.5)$	$\beta_4 = -0.021$	$\beta_{101} = 0.04$
	$X_5 \sim BIN(1, 0.2)$	$X_{17} \sim N(65, 1)$	$\beta_5 = -0.05$	$\beta_{102} = -0.018$
	$X_6 \sim Mult(4, 0.2, 0.3, 0.4, 0.1)$		$\beta_{61} = -0.07$	$\beta_{11} = -0.15$
	$X_7 \sim Mult(4, 0.5, 0.1, 0.2, 0.2)$		$\beta_{62} = 0.04$	$\beta_{12} = -0.08$
	$X_8 \sim Mult(3, 0.3, 0.1, 0.6)$		$\beta_{63} = -0.08$	$\beta_{13} = -0.01$
	$X_9 \sim Mult(3, 0.2, 0.4, 0.4)$		$\beta_{71} = 0.015$	$\beta_{14} = -0.02$
	$X_{10} \sim Mult(3, 0.7, 0.2, 0.1)$		$\beta_{72} = 0.01$	$\beta_{15} = -0.075$
	$X_{11} \sim N(40,3)$		$\beta_{73} = -0.03$	$\beta_{16} = 0$
	$X_{12} \sim N(25,2)$		$\beta_{81} = -0.028$	$\beta_{17} = 0$

B. Data simulasi untuk SURLS-SVM bersifat non-parametrik pada *survival time* yang dibangkitkan. Menurut Kriner (2007) untuk membangkitkan data non-parametrik dapat digunakan interaksi antar variabel prediktor.

Dalam penelitian ini variabel interaksi tidak hanya digunakan sebagai pembentuk *survival time* yang berpola non-linier tetapi juga digunakan untuk melihat bagaimana performansi SURLS-SVM dalam mendeteksi adanya interaksi, karena dalam data riil tidak selalu diketahui apakah ada variabel prediktor yang berinteraksi atau tidak. Dalam bukunya, Kleinbaum dan Klein (2012) membahas mengenai interaksi antara variabel prediktor yang berdistribusi diskrit (status pengobatan) dan variabel yang berdistribusi kontinu (jumlah WBC) dan dinyatakan bahwa interaksinya adalah dengan mengalikan dua variabel tersebut. Dalam penelitian ini, *survival time* yang memuat interaksi dibangkitkan dengan persamaan berikut:

$$T = \left(-\frac{\log(U)}{\lambda \exp(\boldsymbol{\beta}_{full}^T \mathbf{x}_{full})} \right)^{1/\nu},$$

dengan $\boldsymbol{\beta}_{full}^T$ dan \mathbf{x}_{full} berturut-turut adalah vektor koefisien dan prediktor pada Tabel 3.1 dan Tabel 3.2. Skenario untuk prediktor dan koefisien interaksi dirangkum pada Tabel 3.2.

Tabel 3.2 Variabel Prediktor yang Diinteraksikan

Jumlah Prediktor sesuai Tabel 3.1	Prediktor Tambahan	Nilai Parameter
$p = 6$ dan $p = 8$	$X_9 = X_1X_5$ $X_{10} = X_5X_6$	$\beta_9 = 0.1$ $\beta_{10} = -0.001$
$p = 15$ dan $p = 17$	$X_{18} = X_1X_{15}$ $X_{19} = X_1X_{12}$	$\beta_{18} = -0.0001$ $\beta_{19} = 0.25$

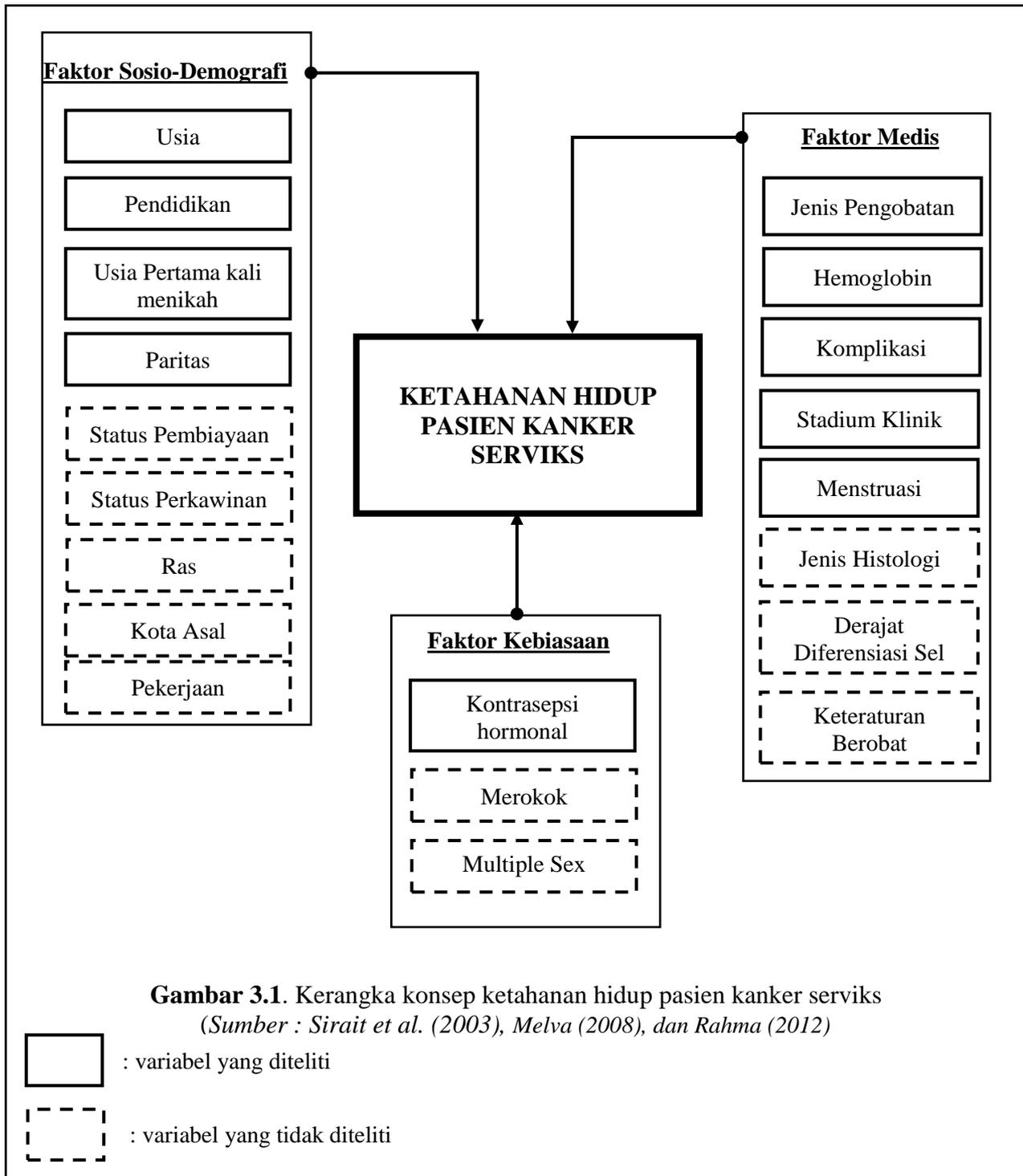
3.2.2 Kerangka Konsep untuk Studi Kasus Pasien Kanker Serviks

Penderita kanker serviks setiap tahun terus meningkat, berdasarkan data dari Kemenkes (2015) penderita kanker serviks pada tahun 2015 berkisar 90-100 kasus per 10.000 penduduk dan setiap tahun terjadi 40 ribu kasus kanker serviks. Penyebab utama kanker serviks diketahui adalah infeksi virus HPV (*Human Papiloma Virus*) sub tipe onkogenik, terutama sub tipe 16 dan 18. Proses terjadinya

kanker serviks dimulai dengan lesi prakanker dan setelah berahun-tahun baru menjadi invasif. Tingkat keparahan penderita kanker serviks berkaitan dengan stadium. Stadium kanker serviks didasarkan atas pemeriksaan klinik, stadium klinik tidak berubah bila kemudian terdapat penemuan baru. Tingkatan dari stadium yaitu 0, IA, IB, IIA, IIB, IIIA, IIIB, IV A, dan IV B. Jenis pengobatan disesuaikan dengan tingkat stadium penyakit yaitu operasi, radiasi dan kemoterapi. Penderita kanker serviks yang akan menjalani jenis pengobatan kemoterapi sebelumnya melakukan pemeriksaan laboratorium yang terdiri dari.

- darah tepi (Hb \geq 10g%, leukosit \geq 3.000/m³, trombosit \geq 100.000/mm³),
- fungsi hati (SGOT <27 U/L, SGPT < 36 U/L),
- ginjal (ureum < 50 mg/d L, kreatinin 0,60-1,20 mg/d L, CCT >68mL/menit).

Berdasarkan penelitian dari sudut pandang medis yang dilakukan Sirait *et al.* (2003), Melva (2008), dan Rahma (2012) selain disebabkan oleh HPV, kanker serviks juga disebabkan oleh menikah di usia yang terlalu muda yaitu di bawah 16 tahun, tingginya persalinan yang dialami, golongan sosial ekonomi rendah, aktivitas seksual sering berganti pasangan, merokok, dan penggunaan kontrasepsi hormonal. Dari tiga penelitian di atas, maka variabel yang mempengaruhi ketahanan hidup pasien kanker serviks dalam penelitian ini meliputi usia, stadium, status komplikasi, penyakit penyerta, status kemoterapi, usia saat menikah, frekuensi persalinan, status penggunaan alat kontrasepsi, dan jenis pengobatan transfusi PRC. Berikut ini adalah kerangka konsep yang digunakan dalam penelitian.



3.2.3 Sumber Data dan Variabel Penelitian

Sebagai perbandingan, pada penelitian ini juga digunakan data riil yaitu data pasien penderita kanker serviks di RS dr. Soetomo Surabaya pada tahun 2014 sampai dengan 2016 sebanyak 412 pasien. Kriteria inklusi yang digunakan dalam penelitian ini adalah sebagai berikut :

- pasien kanker serviks,
- jenis kelamin perempuan,
- memiliki data rekam medik yang lengkap. Kelengkapan yang dimaksud adalah pasien tersebut memiliki data yang sesuai dengan variabel yang akan digunakan dalam penelitian.

Variabel penelitian yang digunakan dalam penelitian ini adalah sebagai berikut.

A. Variabel Respon

Variabel respon terdiri dari *survival time* (T) dan status tersensor (δ). Variabel T dihitung saat pasien masuk dan menjalani rawat inap di RS dr. Soetomo Surabaya hingga pasien dinyatakan meninggal, pindah ke rumah sakit lain, atau berhenti melakukan pengobatan yang dinyatakan dalam satuan hari. Variabel δ merupakan status pasien, bernilai satu jika sebelum penelitian berakhir pasien dinyatakan meninggal, dan bernilai nol jika pasien tersensor. Penyebab tersensor adalah pasien belum meninggal hingga penelitian selesai, pasien tidak dapat diamati lagi karena pindah rumah sakit, berhenti menjalani pengobatan, atau meninggal yang tidak disebabkan oleh kanker serviks selama penelitian. Variabel respon dalam penelitian ini adalah sebagai berikut.

Tabel 3.3 Variabel Respon

Variabel	Keterangan	Skala
Variabel Respon <i>Survival Time</i> (T)	<i>Survival Time</i> pasien kanker serviks	Rasio
Status (δ)	Status terjadi <i>event</i> atau tidak 1 : mengalami <i>event</i> (meninggal) 0 : tidak mengalami <i>event</i> (tersensor)	Nominal

B. Variabel Prediktor

Variabel prediktor yang digunakan adalah variabel yang memberikan pengaruh terhadap waktu ketahanan hidup pasien kanker serviks. Variabel prediktor yang digunakan dalam penelitian ini adalah sebagai berikut.

Tabel 3.4 Variabel Prediktor Pasien Kanker Serviks

Variabel	Keterangan	Skala
P_1	Usia Pasien	Rasio
P_2	Status Komplikasi : 0: Tidak ada 1: Ada	Nominal
P_3	Status anemia 0: Tidak anemia 1: Anemia	Nominal
P_4	Jenis Pengobatan 0: Kemoterapi 1: Transfusi PRC 2: Kemoterapi dan transfusi PRC 3: Lainnya (operasi atau gabungan ketiganya)	Nominal
P_5	Stadium 0 : stadium 1 1 : stadium 2 2 : stadium 3 3 : stadium 4	Ordinal
P_6	Usia saat menikah pertama	Rasio
P_7	Usia saat menstruasi pertama	Rasio
P_8	Siklus menstruasi	Rasio
P_9	Lama menstruasi	Rasio
P_{10}	Paritas	Rasio
P_{11}	Jenis KB 0: tidak KB 1: KB suntik 2: KB Pil 3: KB IUD	Nominal
P_{12}	Tingkat pendidikan 0: SD 1: SMP 2: SMA 3: Pendidikan tinggi (setelah SMA)	Ordinal

Definisi operasional untuk masing-masing variabel prediktor adalah sebagai berikut.

- usia pasien: usia pasien saat pertama kali melakukan pengobatan,

- status komplikasi: memberikan informasi ada atau tidaknya komplikasi penyakit lain yang diderita pasien kanker serviks,
- status anemia: memberikan informasi apakah pasien mengalami anemia atau tidak,
- jenis pengobatan: memberikan informasi jenis pengobatan yang dilakukan pasien,
- stadium: memberikan informasi stadium kanker dari pasien kanker serviks,
- usia saat menikah pertama: usia pasien kanker serviks saat menikah pertama kali,
- usia saat menstruasi pertama: usia pasien kanker serviks saat menstruasi pertama kalinya,
- siklus menstruasi: jarak mulainya menstruasi dengan menstruasi berikutnya,
- lama menstruasi: lama menstruasi dalam satu periode menstruasi,
- paritas: frekuensi melahirkan hidup,
- jenis KB: memberikan informasi mengenai jenis KB yang dilakukan,
- tingkat pendidikan: memberikan informasi mengenai pendidikan.

Perbedaan data pasien kanker serviks yang digunakan pada penelitian ini dengan data pada penelitian Afifah dan Purnami (2016) terletak pada jumlah variabel independen. Berdasarkan literatur dalam bidang kesehatan seperti pada Gambar 3.1, ada beberapa variabel independen yang mempengaruhi ketahanan hidup pasien kanker serviks tetapi belum teramati karena terbatasnya data yang tersedia. Variabel-variabel baru yang membedakan adalah paritas, lama menstruasi, usia pertama kali menstruasi, usia saat menikah pertama kali, siklus menstruasi, lama menstruasi, jenis kontrasepsi, dan tingkat pendidikan.

3.2.4 Langkah Analisis Data

Tahapan yang dilakukan dalam analisis data dijelaskan dalam uraian berikut:

1. Membangkitkan data yang digunakan untuk simulasi dengan langkah-langkah sebagai berikut.
 - a. Menentukan ukuran data yang dibangkitkan. Dalam penelitian ini banyak data yang dibangkitkan adalah $n_1 = 100$, $n_2 = 1.000$, dan $n_3 = 5.000$.

- b. Membangkitkan data untuk variabel prediktor menggunakan distribusi yang sesuai dengan skenario simulasi pada sub bab 3.2.1.
 - c. Membangkitkan data untuk variabel respon dibagi menjadi dua jenis, yaitu data yang bersifat parametrik dan data yang bersifat non-parametrik. Untuk data dengan pendekatan parametrik menggunakan skenario poin A pada sub bab 3.2.1 dan menggunakan persamaan (2.32). Untuk data dengan pendekatan non-parametrik menggunakan skenario B pada sub bab 3.2.1.
 - d. Membangkitkan status tersensor dengan distribusi binomial bersesuaian dengan persentase tersensor. Misal persentase tersensor 10%, maka distribusi yang digunakan untuk membangkitkan status tersensor adalah $BIN(1,0.9)$ dan banyaknya bangkitan menyesuaikan *sample size* (n) yang digunakan.
2. Melakukan analisis deskriptif untuk data riil yaitu data pasien kanker serviks yang menjalani rawat inap di RS dr.Soetomo Surabaya, dan membuat variabel *dummy* untuk variabel-variabel yang memiliki skala ordinal dan nominal, misalkan variabel jenis pengobatan, dimana variabel *dummy* yang digunakan ada tiga karena pada data ada empat kategori pengobatan.
 3. Melakukan analisis data simulasi dan data riil dengan langkah-langkah sebagai berikut :
 - a. Mengoptimasi parameter γ dan σ^2 dengan metode *grid search*.
 - b. Berdasarkan nilai parameter yang telah dioptimasi, selanjutnya digunakan untuk membangun model SURLSSVM seperti pada persamaan (2.27) dengan menggunakan fungsi kernel Gaussian.
 - c. Menghitung nilai prediksi dari *prognostic index* SURLSSVM seperti pada persamaan (2.31)
 - d. Melakukan analisis data *survival* dengan model regresi Cox PHM dan menghitung *prognostic index* dari Cox PHM.
 4. Menentukan model yang terbaik dengan menghitung performansi untuk Cox PHM dan SURLS-SVM.
 - a. Menghitung nilai c-index untuk model SURLSSVM dengan menggunakan persamaan (2.32).

- b. Menghitung nilai log-rank test berdasarkan prediksi *prognostic index*. *Prognostic index* dibagi menjadi dua kelompok yaitu *high risk* dan *low risk*. Pemecahan ini dapat dilakukan berdasarkan nilai mean atau nilai median dari *prognostic index*. Untuk menguji apakah model mampu memisahkan pasien *high risk* dan *low risk* digunakan uji Log Rank seperti pada persamaan (2.10).
- c. Menentukan *hazard ratio* yang diperoleh dengan menormalisasi *prognostic index*. Untuk menormalisasi *prognostic index* ke skala nol sampai dengan satu digunakan persamaan berikut

$$\hat{u}_{new} = \left(\frac{\hat{u}_{old} - \hat{u}_{min}}{\hat{u}_{max} - \hat{u}_{min}} \right) (BA - BB) + BB \quad (3.2)$$

dengan :

\hat{u}_{new} : *prognostic index* setelah dinormalisasi

\hat{u}_{old} : *prognostic index* sebelum dinormalisasi

\hat{u}_{min} : nilai *prognostic index* yang paling minimum sebelum dinormalisasi

\hat{u}_{max} : nilai *prognostic index* yang paling maksimum sebelum dinormalisasi

BA : batas atas (dalam penelitian ini $BA = 1$)

BB : batas bawah (dalam penelitian ini $BB=0$).

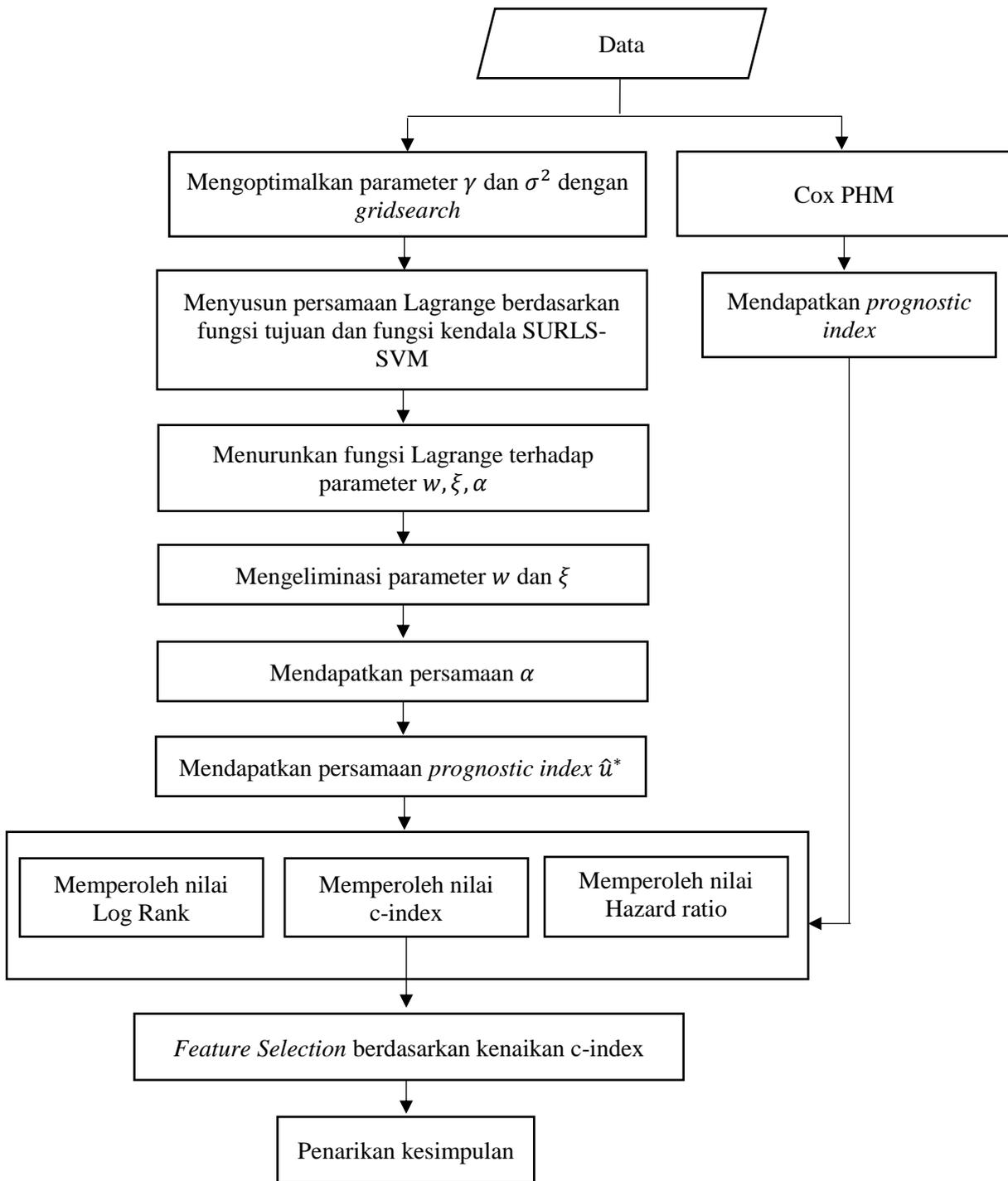
Setelah dilakukan normalisasi selanjutnya diregresikan menggunakan metode regresi Cox dengan variabel prediktornya adalah *prognostic index* yang baru, dan selanjutnya dihitung nilai dari *hazard ratio*.

5. Menentukan variabel yang signifikan.

Model SURLS-SVM tidak dapat digunakan secara langsung untuk mengetahui variabel prediktor yang signifikan sebagaimana yang bisa diperoleh dari Cox PHM. Oleh karena itu dalam penelitian ini ditambahkan metode *feature selection* untuk mendapatkan variabel yang berpengaruh. Ukuran yang digunakan untuk memilih variabel yang masuk dalam model adalah c-index.

6. Penarikan kesimpulan berdasarkan hasil perbandingan antara Cox PHM dan model SURLS-SVM.

Secara umum, langkah-langkah penelitian ditunjukkan pada Gambar 3.2.



Gambar 3.2 Langkah Analisis

3.2.5 Struktur Data Penelitian

Data yang digunakan dalam penelitian ini adalah data *survival* yang tidak berulang dan memiliki variabel prediktor yang bersifat *time-independent*. Struktur datanya adalah sebagai berikut

Tabel 3.5 Struktur Data *Survival*

Individu (i)	T_i	δ_i	X_{i1}	X_{i2}	...	X_{i12}
1	T_1	δ_1	X_{11}	X_{12}	...	X_{112}
2	T_2	δ_2	X_{21}	X_{22}	...	X_{212}
\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
n	T_n	δ_n	X_{n1}	X_{n2}	...	X_{n12}

Keterangan:

T_i : *survival time*

δ_i : status tersensor

X_1, X_2, \dots, X_{12} : variabel prediktor

(Halaman ini sengaja dikosongkan)

BAB 4

HASIL DAN PEMBAHASAN

Pada bab ini diuraikan langkah-langkah untuk mencapai tujuan dalam penelitian. Secara garis besar adalah melakukan kajian teoritis untuk model SURLS-SVM kemudian diterapkan pada data riil dan data simulasi.

4.1 Kajian Teoritis Model SURLS-SVM

Pada Cox PHM untuk menuliskan model yang terbentuk digunakan *hazard function*, tidak secara langsung menuliskan persamaannya dalam *survival time* karena *survival time* sulit untuk diprediksi. Selain menggunakan *hazard function*, dapat juga digunakan *prognostic index* yang nantinya akan diukur kesesuaian ranking antara *survival time* dan *prognostic index*.

Pada sub-bab ini diuraikan langkah-langkah kajian teoritis dari model SURLS-SVM untuk mendapatkan *prognostic index* seperti yang dituliskan pada (2.31). Uraian untuk mendapatkan *prognostic index* adalah sebagai berikut:

Untuk mendapatkan persamaan *prognostic index* diawali dengan menurunkan persamaan dari fungsi tujuan dan fungsi kendala model SURLS-SVM. Fungsi tujuan dan fungsi kendala SURLS-SVM yang sudah dituliskan pada persamaan (2.27) adalah sebagai berikut:

Fungsi tujuan

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

Fungsi kendala

$$\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) = 1 + \xi_{ij}, \quad \forall i, j = 1, 2, \dots, n$$

Untuk mendapatkan nilai optimal dari fungsi tujuan, dibentuk persamaan Lagrange seperti pada persamaan (2.28) yang menggabungkan antara fungsi tujuan dan fungsi kendala sebagai berikut:

$$L(\mathbf{w}, \xi; \boldsymbol{\alpha}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \frac{\gamma}{2} \sum_{i=1}^n \sum_{i < j}^n v_{ij} \xi_{ij}^2 - \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_j) - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - 1 - \xi_{ij})$$

dimana α_{ij} adalah pengali Lagrange. Untuk mendapatkan penyelesaian optimal, persamaan Lagrange diturunkan terhadap masing-masing parameter yaitu \mathbf{w}, ξ, α , sebagai berikut:

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \rightarrow \mathbf{w} - \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) = 0 \rightarrow \mathbf{w} = \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \\ \frac{\partial L}{\partial \xi_{ij}} = 0 \rightarrow \gamma \sum_{i=1}^n \sum_{i < j}^n v_{ij} \xi_{ij} - \sum_{i=1}^n \sum_{i < j}^n (-\alpha_{ij}) = 0 \rightarrow \alpha_{ij} = -\gamma v_{ij} \xi_{ij} \\ \frac{\partial L}{\partial \alpha_{ij}} = 0 \rightarrow -\sum_{i=1}^n \sum_{i < j}^n \mathbf{w}^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) - 1 - \xi_{ij} = 0 \rightarrow \sum_{i=1}^n \sum_{i < j}^n 1 + \xi_{ij} - \mathbf{w}^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) = 0 \end{cases} \quad (4.1)$$

Untuk menyederhanakan persamaan (4.1), dilakukan eliminasi dan substitusi. Sebelum melakukan eliminasi, dilakukan substitusi untuk parameter \mathbf{w} dan ξ_{ij} sehingga persamaan untuk parameter \mathbf{w} dan ξ_{ij} diubah menjadi persamaan berikut

$$\mathbf{w}^T = \left[\sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right]^T \text{ dan } \xi_{ij} = -\frac{\alpha_{ij}}{\gamma v_{ij}} \text{ dan disusun persamaan baru dengan}$$

substitusi dan eliminasi parameter \mathbf{w} dan ξ_{ij} sebagai berikut

$$\begin{aligned} \sum_{i=1}^n \sum_{i < j}^n 1 + \xi_{ij} - \mathbf{w}^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) &= 0 \\ \Leftrightarrow \sum_{i=1}^n \sum_{i < j}^n 1 - \sum_{i=1}^n \sum_{i < j}^n \frac{\alpha_{ij}}{\gamma v_{ij}} - \left\{ \left[\sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right]^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right\} &= 0 \end{aligned} \quad (4.2)$$

Persamaan (4.2) dapat disederhanakan dengan mengalikan masing-masing suku dengan γv_{ij} sehingga diperoleh persamaan sebagai berikut:

$$\begin{aligned} \gamma \sum_{i=1}^n \sum_{i < j}^n v_{ij} - \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} - \gamma \sum_{i=1}^n \sum_{i < j}^n v_{ij} \left\{ \left[\sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right]^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right\} &= 0 \\ \Leftrightarrow \gamma \sum_{i=1}^n \sum_{i < j}^n v_{ij} = \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} + \gamma \sum_{i=1}^n \sum_{i < j}^n v_{ij} \left\{ \left[\sum_{i=1}^n \sum_{i < j}^n \alpha_{ij} (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right]^T (\varphi(\mathbf{x}_j) - \varphi(\mathbf{x}_i)) \right\} & \quad (4.3) \end{aligned}$$

v_{ij} merupakan indikator apakah dua individu dapat dibandingkan atau tidak yang didefinisikan pada persamaan (2.26) dan menginterpretasikan urutan *survival time* dengan mempertimbangkan status tersensor. Dalam hal ini nilai v_{ij} yang diambil

hanya untuk individu yang dapat dibandingkan atau nilai v_{ij} sama dengan 1.

Penjabaran persamaan (4.3) jika dibuat dalam suatu matriks adalah sebagai berikut

$$\gamma \sum_{i=1}^n \sum_{i<j}^n v_{ij} = \gamma(v_{12} + v_{13} + \dots + v_{(n_c-1)n_c}) = \gamma \mathbf{1}_{(n_c \times 1)} \quad (4.4)$$

n_c adalah banyaknya individu yang dapat dibandingkan.

$$\sum_{i=1}^n \sum_{i<j}^n \alpha_{ij} = \alpha_{11} + \alpha_{12} + \dots + \alpha_{(n-1)n} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_{12} \\ \alpha_{13} \\ \vdots \\ \alpha_{(n-2)n} \\ \alpha_{(n-1)n} \end{bmatrix} = \mathbf{I}_{(n_c \times n_c)} \boldsymbol{\alpha}_{(n_c \times 1)} \quad (4.5)$$

Ruas kanan merupakan penjumlahan yang melibatkan matriks identitas sehingga ukuran matriks adalah persegi. Agar dapat dijumlahkan ukuran matriks harus sama, sehingga

$$\gamma \sum_{i=1}^n \sum_{i<j}^n v_{ij} = \gamma(v_{12} + v_{13} + \dots + v_{(n_c-1)n_c}) = \gamma \begin{bmatrix} v_{12} & 0 & \dots & 0 \\ 0 & v_{13} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & v_{(n_c-1)n_c} \end{bmatrix}$$

$$\Leftrightarrow \gamma \sum_{i=1}^n \sum_{i<j}^n v_{ij} = \gamma \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} = \gamma \mathbf{I}_{(n_c \times n_c)} \quad (4.6)$$

Berdasarkan hasil persamaan (4.4) - (4.6), persamaan (4.3) dapat dinyatakan dalam bentuk matriks sebagai berikut:

$$\gamma \sum_{i=1}^n \sum_{i<j}^n v_{ij} = \sum_{i=1}^n \sum_{i<j}^n \alpha_{ij} + \gamma \sum_{i=1}^n \sum_{i<j}^n v_{ij} \left\{ \left[\sum_{i=1}^n \sum_{i<j}^n \alpha_{ij} (\varphi(x_j) - \varphi(x_i)) \right]^T (\varphi(x_j) - \varphi(x_i)) \right\}$$

Selanjutnya berdasarkan matriks \mathbf{D} pada Bab 2 yang dikalikan dengan \mathbf{X} yaitu vektor variabel prediktor, maka karena dalam SURLS-SVM menggunakan kernel, matriks \mathbf{D} akan dikalikan dengan kernel. Parameter kernel yaitu $\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)$ memiliki dua bentuk $(\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j))$ dan $(\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j))^T$, sehingga ada dua matriks \mathbf{D} yang menyertai kernel. Sehingga didapatkan persamaan berikut:

$$\begin{aligned}
&\Leftrightarrow \gamma \mathbf{1} = \mathbf{I} \boldsymbol{\alpha} + \gamma \mathbf{I} (\boldsymbol{\alpha}^T \mathbf{D} \mathbf{K}^T \mathbf{D}^T)^T \\
&\Leftrightarrow \gamma \mathbf{1} = \mathbf{I} \boldsymbol{\alpha} + \gamma \mathbf{D} \mathbf{K} \mathbf{D}^T \boldsymbol{\alpha} \\
&\Leftrightarrow \gamma \mathbf{D} \mathbf{K} \mathbf{D}^T \boldsymbol{\alpha} + \mathbf{I} \boldsymbol{\alpha} = \gamma \mathbf{1} \\
&\Leftrightarrow (\gamma \mathbf{D} \mathbf{K} \mathbf{D}^T + \mathbf{I}) \boldsymbol{\alpha} = \gamma \mathbf{1} \\
&\Leftrightarrow \boldsymbol{\alpha} = (\gamma \mathbf{D} \mathbf{K} \mathbf{D}^T + \mathbf{I})^{-1} \gamma \mathbf{1} \tag{4.7}
\end{aligned}$$

Setelah mendapatkan persamaan untuk $\boldsymbol{\alpha}$, selanjutnya ditentukan persamaan untuk *prognostic index* yang ditunjukkan dalam persamaan (4.8).

$$\begin{aligned}
\hat{\mathbf{u}}^* &= \sum_{i=1}^n \sum_{i < j}^n [\hat{\alpha}_{ij} (\boldsymbol{\varphi}(\mathbf{x}_i) - \boldsymbol{\varphi}(\mathbf{x}_j))]^T \boldsymbol{\varphi}(\mathbf{x}^*) \\
&= \sum_{i=1}^n \sum_{i < j}^n \hat{\alpha}_{ij}^T [\boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}^*) - \boldsymbol{\varphi}(\mathbf{x}_j)^T \boldsymbol{\varphi}(\mathbf{x}^*)] \\
&= \hat{\boldsymbol{\alpha}}^T \mathbf{D} \mathbf{K}_n (\mathbf{x}^*) \tag{4.8}
\end{aligned}$$

Ukuran untuk masing-masing matriks adalah sebagai berikut:

$$\mathbf{D}_{(n_c \times n)}, \mathbf{K}_{(n \times n)}, \mathbf{I}_{(n_c \times n_c)}, \mathbf{1}_{(n_c \times 1)}, \boldsymbol{\alpha}_{(n_c \times 1)}.$$

Elemen pada matriks \mathbf{D} bergantung pada status tersensor dan ranking dari masing-masing *survival time*. Contoh menentukan elemen pada matriks \mathbf{D} adalah sebagai berikut:

Misalkan terdapat 5 data *survival* terdiri dari *survival time* dan status tersensor yang ditampilkan pada Tabel 4.1.

Tabel 4.1. Contoh Data *Survival*

Pasien ke-	<i>Survival Time</i>	Status		X_1	X_2
		1: Terjadi <i>event</i>	0: Tersensor		
1	10	0		0,5	1
2	3	1		0,3	0
3	5	0		0,7	0
4	2	0		1	1
5	15	1		2	0

Langkah 1: Melakukan Inisiasi Vektor untuk *Survival Time* dan Status

$$\mathbf{T} = \begin{pmatrix} 10 \\ 3 \\ 5 \\ 2 \\ 15 \end{pmatrix}, \boldsymbol{\delta} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Langkah 2: Mengurutkan *survival time* dari terkecil ke terbesar. Status tersensor mengikuti urutan *survival time*. Dengan menggunakan data pada Tabel 4.1, diperoleh urutannya sebagai berikut:

$$\mathbf{T}_{\text{new}} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 10 \\ 15 \end{pmatrix}, \delta_{\text{new}} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{\pi} = \begin{pmatrix} 4 \\ 2 \\ 3 \\ 1 \\ 5 \end{pmatrix}$$

dengan $\boldsymbol{\pi}$ adalah urutan asli dari *survival time*.

Langkah 3: Menentukan inisialisasi nilai

a. Mencari urutan dari *event* yang pertama kali terjadi (e_1)

Berdasarkan contoh data yang digunakan, *event* pertama kali terjadi pada urutan kedua karena status tidak tersensor pertama kali terjadi pada *survival time* (T) urutan kedua, sehingga inisialisasi awal untuk $e_1 = 2$.

b. Menentukan ukuran \mathbf{D}

Banyaknya kolom (n) untuk matriks \mathbf{D} adalah sama dengan banyaknya *survival time*, dimana pada data yang digunakan sebagai contoh nilainya adalah $n = 5$. Banyaknya baris (n_c) untuk matriks \mathbf{D} adalah selisih antara banyaknya *survival time* (n) dengan letak dari *event* pertama (e_1). Karena pada contoh nilai n dan e_1 berturut-turut adalah 5 dan 2, maka banyaknya baris untuk matriks \mathbf{D} adalah 3, sehingga ukuran untuk matriks \mathbf{D} adalah $\mathbf{D}_{(3 \times 5)}$.

c. Menentukan inisialisasi untuk matriks \mathbf{D}

Inisialisasi untuk matriks \mathbf{D} adalah matriks nol yang berukuran sama dengan matriks \mathbf{D} , sehingga inisialisasi untuk matriks \mathbf{D} adalah sebagai berikut:

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

Langkah 4: Iterasi elemen dari matriks \mathbf{D}

Iterasi dilakukan per baris, karena pada data contoh terdapat tiga baris maka iterasi dilakukan tiga kali atau $i = 1, 2, 3$. Untuk iterasi pertama ($i = 1$) dilakukan penggantian nilai di beberapa kolom yang terletak pada baris satu. Penggantian nilai

-1 dan 1 bergantung pada posisi T. Untuk iterasi 1 membandingkan T=3 dan T=5, karena *event* pertama terjadi pada saat T=3. a_i digunakan untuk kolom yang diganti dengan -1 dan b_i digunakan untuk kolom yang diganti dengan 1. Untuk baris pertama ($i = 1$), $a_1 = 2$ karena *event* terjadi pada *survival time* urutan kedua sehingga $D[1,2] = -1$. Selanjutnya nilai $b_1 = e_1 + i = 3$, sehingga $D[1,3] = 1$, dan diperoleh baris 1 untuk matriks **D** adalah $(0 \quad -1 \quad 1 \quad 0 \quad 0)$.

Selanjutnya untuk iterasi 2 yang dilakukan pada baris kedua adalah membandingkan $T=3$ dan $T=10$, $a_2 = 2$ dimana nilainya sama dengan a_1 karena *event* pada urutan 1 sampai 3 tetap terletak pada 2. Status yang dilihat untuk iterasi kedua hanya sampai 3 atau satu urutan sebelum $T=10$ yang terletak pada urutan ke-4, dan nilai $b_2 = e_1 + 2 = 4$ sehingga $D[2,2] = -1$ dan $D[2,4] = 1$ atau matriks **D** pada baris kedua adalah $(0 \quad -1 \quad 0 \quad 1 \quad 0)$.

Untuk iterasi 3 dilakukan pada baris ketiga yang membandingkan $T=3$ dengan $T=15$. $T=3$ selalu dibandingkan dengan T sesudahnya karena pada T sesudah $T=3$ tidak ada lagi yang mengalami *event* dan baru terjadi lagi pada $T=15$. Pada iterasi 3, nilai $a_3 = 2$ dan $b_3 = e_1 + i = 5$ sehingga $D[3,2] = -1$ dan $D[3,5] = 1$ atau matriks **D** pada baris ketiga adalah $(0 \quad -1 \quad 0 \quad 0 \quad 1)$ sehingga

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

Setelah matriks **D** terbentuk selanjutnya dibentuk matriks kernel (**K**) yang merepresentasikan variabel prediktor. Dalam penelitian ini kernel yang digunakan adalah RBF kernel yang dihitung dengan menggunakan persamaan $\mathbf{K}(\mathbf{a}_i, \mathbf{a}_j) = \exp\left(-\frac{\|\mathbf{a}_i - \mathbf{a}_j\|^2}{2\sigma^2}\right)$. Misalkan akan dihitung kernel antara pasien ke-1 dan pasien ke-2 yang memiliki variabel prediktor berturut-turut $\mathbf{a}_1 = \begin{pmatrix} 0,5 \\ 1 \end{pmatrix}$, $\mathbf{a}_2 = \begin{pmatrix} 0,3 \\ 0 \end{pmatrix}$ dengan parameter kernel $\sigma^2 = 0,2$, maka untuk menghitung nilai kernel pasien 1 dan 2 digunakan persamaan berikut

$$\mathbf{K}(\mathbf{a}_i, \mathbf{a}_j) = \exp\left(-\left(\frac{((0,5^2 + 1^2) - 2((0,5 \times 0,3) + (1 \times 0)) + (0,3^2 + 0))}{2(0,2)}\right)\right) = 0,0743.$$

Matriks \mathbf{K} selengkapnya untuk semua pasien adalah sebagai berikut:

$$\mathbf{K}(\mathbf{a}_i, \mathbf{a}_j) = \begin{pmatrix} 1 & 0,0743 & 0,0743 & 0,5353 & 0,0003 \\ 0,0743 & 1 & 0,6703 & 0,0241 & 0,0007 \\ 0,0743 & 0,6703 & 1 & 0,0655 & 0,0146 \\ 0,5353 & 0,0241 & 0,0655 & 1 & 0,00067 \\ 0,0003 & 0,0007 & 0,0146 & 0,00067 & 1 \end{pmatrix}$$

Selanjutnya matriks \mathbf{K} diurutkan sesuai dengan urutan *survival time* (T) dan didapatkan matriks \mathbf{K} sebagai berikut:

$$\mathbf{K}(\mathbf{a}_i, \mathbf{a}_j) = \begin{pmatrix} 0,5353 & 0,0241 & 0,0655 & 1 & 0,00067 \\ 0,0743 & 1 & 0,6703 & 0,0241 & 0,0007 \\ 0,0743 & 0,6703 & 1 & 0,0655 & 0,0146 \\ 1 & 0,0743 & 0,0743 & 0,5353 & 0,0003 \\ 0,0003 & 0,0007 & 0,0146 & 0,00067 & 1 \end{pmatrix}$$

Setelah mendapatkan matriks \mathbf{K} dan \mathbf{D} selanjutnya dapat dihitung nilai dari vektor $\hat{\alpha}$ dengan menggunakan persamaan (4.7). Nilai $\gamma = 0,1$, ukuran untuk matriks $\mathbf{1}$ adalah (3×1) karena perbandingan pasien dapat dilakukan 3 kali. Vektor $\hat{\alpha}$ yang terbentuk berukuran (3×1) sebagai berikut

$$\hat{\alpha} = \begin{pmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0,5353 & 0,0241 & 0,0655 & 1 & 0,00067 \\ 0,0743 & 1 & 0,6703 & 0,0241 & 0,0007 \\ 0,0743 & 0,6703 & 1 & 0,0655 & 0,0146 \\ 1 & 0,0743 & 0,0743 & 0,5353 & 0,0003 \\ 0,0003 & 0,0007 & 0,0146 & 0,00067 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix}^T + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$\times 0,1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\hat{\alpha} = \begin{pmatrix} 0,0887 \\ 0,0789 \\ 0,0744 \end{pmatrix}$$

Sehingga untuk menghitung *prognostic index* berdasarkan persamaan (4.8) digunakan nilai-nilai berikut:

$$\hat{\mathbf{u}} = (0,0887 \quad 0,0789 \quad 0,0744) \begin{pmatrix} 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0,5353 & 0,0241 & 0,0655 & 1 & 0,00067 \\ 0,0743 & 1 & 0,6703 & 0,0241 & 0,0007 \\ 0,0743 & 0,6703 & 1 & 0,0655 & 0,0146 \\ 1 & 0,0743 & 0,0743 & 0,5353 & 0,0003 \\ 0,0003 & 0,0007 & 0,0146 & 0,00067 & 1 \end{pmatrix}$$

$$\hat{\mathbf{u}} = (0,0675 \quad -0,1765 \quad -0,0665 \quad 0,0427 \quad 0,0755)$$

Berdasarkan hasil yang diperoleh, maka *prognostic index* untuk pasien pertama adalah 0,0427 karena pasien ke-4 memiliki $T=10$ yang terletak pada urutan ke-4. Berdasarkan *prognostic index*, nilai mediannya adalah 0,0427, sehingga pasien yang memiliki prognostic kurang dari 0,0427 masuk dalam kategori *low risk* sedangkan yang memiliki prognostic lebih dari 0,0427 masuk dalam kategori *high risk*.

4.2 Penerapan SURLS-SVM dan Cox PHM untuk Data Simulasi

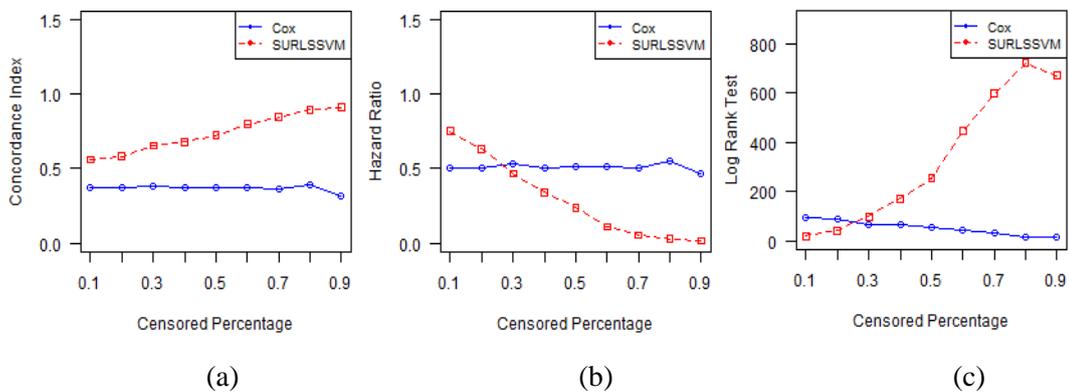
Cox PHM merupakan model semi-parametrik yang paling sering digunakan untuk analisis data *survival*, dimana pada Cox PHM yang dicari adalah model dari fungsi hazard. Karena Cox PHM memiliki asumsi yang harus dipenuhi yaitu Proporsional Hazard atau nilai hazard tetap untuk setiap waktu, maka diperlukan model lain yang dapat digunakan saat Cox PHM tidak semua variabelnya memenuhi asumsi PH. Salah satu model yang dapat digunakan adalah SURLS-SVM. Jika pada Cox PHM menggunakan fungsi hazard maka pada SURLS-SVM menggunakan *prognostic index* sebagai pengganti fungsi hazard. *Prognostic index* untuk Cox PHM diperoleh dengan mengalikan estimasi parameter dengan prediktornya. Jika pada Cox PHM yang digunakan sebagai input adalah *survival time* dan prediktor dengan mempertimbangkan status tersensor, maka pada SURLS-SVM yang digunakan sebagai input adalah ranking dari *survival time* dengan mempertimbangkan indikator pembanding sebagaimana yang dituliskan pada persamaan (2.26) serta prediktor yang menggunakan kernel RBF sebagai *feature mapping*.

Untuk analisis pertama digunakan data simulasi yang dibangkitkan dengan distribusi seperti yang dijelaskan dalam Bab III, meliputi *survival time*, status tersensor, dan variabel prediktor. Banyaknya variabel prediktor mengikuti skenario yang telah ditentukan dan *sample size* yang dibangkitkan adalah 100, 1000, dan 5000. Dari simulasi yang dilakukan dapat dilihat performansi untuk masing-masing metode yaitu Cox PHM dan SURLS-SVM. Hasil simulasi untuk masing-masing jenis data adalah sebagai berikut:

4.2.1 Performansi SURLS-SVM dan Cox PHM pada Data Linier dan Non-Linier

Data linier diperoleh dengan membangkitkan *survival time* yang berdistribusi Weibull menggunakan persamaan (2.40), sedangkan data non-linier dibangkitkan menggunakan persamaan yang sama dan ditambahkan interaksi antar dua variabel. Variabel-variabel prediktor dibangkitkan sesuai dengan skenario yang telah ditentukan. *Sample size* yang dibangkitkan adalah 100, 1000, dan 5000, dengan jumlah prediktor 6, 8, 15, dan 17. Variasi ini dimaksudkan untuk mengetahui bagaimanakah performansi dari model saat digunakan pada *sample size*, jumlah prediktor, dan persentase tersensor yang berbeda-beda. Data simulasi dianalisa menggunakan Cox PHM dan SURLS-SVM. Untuk model SURLS-SVM digunakan parameter γ dan σ^2 berturut-turut yaitu 0,1 dan 0,5. Performansi model dilihat berdasarkan tiga ukuran, yaitu concordance index (c-index), hazard ratio, dan log-rank. Ketiga ukuran performansi tersebut didapatkan dari *prognostic index* yang merupakan pengganti dari fungsi hazard.

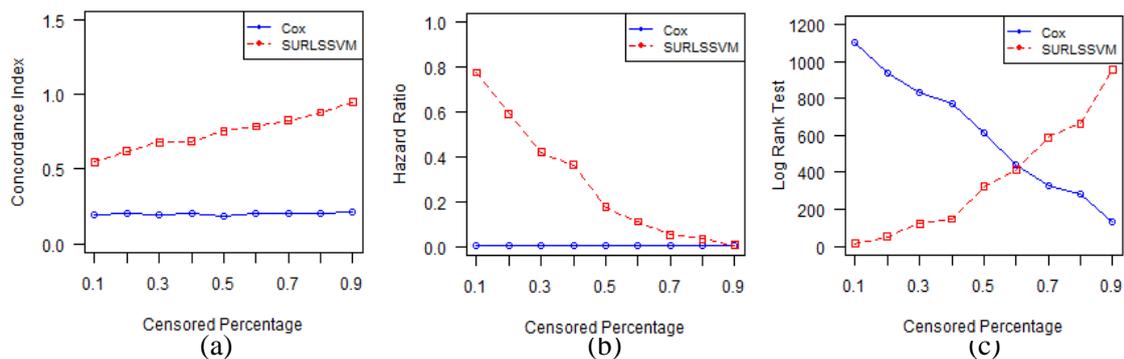
Perbedaan antara data yang berpola linier dan data yang berpola non-linier adalah pada saat membangkitkan *survival time*. Pada data linier tidak ada interaksi antar variabel prediktor, sedangkan pada data non-linier terdapat interaksi antar variabel prediktornya. Performansi pada data linier dan non-linier dilihat pada berbagai persentase tersensor yaitu 0,1; 0,2; ...; 0,9. Hasil dari performansi untuk data linier dengan jumlah variabel prediktor 17 dan *sample size* 1000 pada berbagai persentase tersensor ditampilkan pada Gambar 4.1 berikut:



Gambar 4.1 Performansi Data Simulasi Linier ($n=1000$, $p=17$)
(a)c-index, (b)HR, (c)Log-Rank

Pada Gambar 4.1 (a) dapat dilihat performansi dari model SURLS-SVM selalu lebih unggul daripada model Cox PHM untuk semua persentase tersensor jika dilihat berdasarkan concordance index (c-index). Semakin tinggi persentase tersensor, semakin besar selisih nilai c-index antara Cox PHM dan SURLS-SVM. Hal ini dapat terjadi karena saat persentase tersensor tinggi artinya hanya sedikit objek yang dapat dibandingkan sehingga akan memperkecil tingkat kesalahan ranking (*missranking*). Ukuran performansi kedua yang digunakan adalah hazard ratio (HR). Berdasarkan Gambar 4.1(b), nilai HR pada model SURLS-SVM lebih tinggi daripada Cox PHM pada persentase tersensor rendah (dalam kasus ini di bawah 30%) dan untuk persentase tersensor tinggi Cox PHM memberikan nilai HR yang jauh lebih tinggi. Ukuran performansi ketiga adalah Log-Rank dimana nilai log-rank ini menentukan apakah model mampu memisahkan pasien menjadi grup *low risk* dan *high risk*. Berdasarkan Gambar 4.1(c), model SURLS-SVM memiliki nilai log-rank lebih tinggi dari Cox PHM pada saat persentase tersensor di atas 20%, artinya hanya pada saat persentase tersensor 10% dan 20% Cox PHM memberikan nilai log-rank lebih besar, dan selisih nilai log-rank antara SURLS-SVM dan Cox PHM saat persentase tersensor 10% dan 20% sangat kecil. Sama seperti c-index, SURLS-SVM memiliki nilai log-rank yang semakin besar seiring dengan bertambahnya status tersensor.

Selain menggunakan data yang memiliki pola linier, pada penelitian ini juga digunakan data yang memiliki pola non-linier dimana saat membangkitkan *survival time* (T) terdapat variabel yang saling berinteraksi. Hasil performansi untuk data non-linier yang memiliki 17 variabel prediktor dan *sample size* 1000 pada berbagai persentase tersensor ditampilkan pada Gambar 4.2 berikut:



Gambar 4.2 Performansi Data Simulasi Non-Linier ($n=1000$, $p=17$)
(a)c-index, (b)HR, (c)Log-Rank

Berdasarkan Gambar 4.2, SURLS-SVM selalu lebih unggul dari Cox PHM pada semua persentase tersensor berdasarkan nilai concordance index (c-index) seperti yang ditunjukkan pada Gambar 4.2(a). Sama seperti pada data linier, c-index SURLS-SVM yang tinggi diperoleh saat persentase tersensor tinggi. Ukuran performansi yang kedua adalah hazard ratio (HR) yang ditampilkan pada Gambar 4.2(b). HR dari SURLS-SVM cenderung lebih tinggi dari Cox PHM di hampir semua persentase tersensor, walaupun semakin besar persentase tersensor semakin kecil nilai HR dan selisih nilai HR antara SURLS-SVM dan Cox PHM juga semakin kecil. Nilai HR dari Cox PHM cenderung konstan untuk semua persentase tersensor. Ukuran performansi ketiga adalah nilai log-rank yang ditampilkan pada Gambar 4.2(c). Nilai log-rank untuk SURLS-SVM lebih rendah dari Cox PHM saat persentase tersensor di bawah 60%, dan semakin meningkat persentase tersensor maka semakin meningkat pula nilai log-rank. Berlawanan dengan SURLS-SVM, nilai log-rank Cox PHM justru semakin turun saat persentase tersensor semakin naik sehingga pada saat persentase tersensor di atas 50%, nilai log-rank dari Cox PHM lebih rendah dari SURLS-SVM.

Hasil performansi yang diperoleh antara data yang memiliki pola linier dengan data yang berpola non-linier cenderung berbeda. Jika dilihat berdasarkan nilai c-index yang ditampilkan pada Gambar 4.1(a) dan 4.1(b), SURLS-SVM selalu lebih unggul dari Cox PHM. Namun jika diperhatikan lagi, terdapat perbedaan selisih yang cukup mencolok antara c-index SURLS-SVM dengan Cox PHM pada data berpola linier dan selisih c-index kedua model pada data yang berpola non-linier. Pada data berpola linier, selisih c-index antara SURLS-SVM dan Cox PHM cenderung kecil, tetapi pada data berpola non-linier terdapat selisih yang cukup lebar antara kedua model. Hal ini menunjukkan bahwa c-index memberikan hasil yang lebih baik jika diterapkan pada data yang berpola non-linier. Selain itu, nilai c-index dari Cox PHM juga cenderung lebih konstan pada data yang berpola non-linier, artinya persentase tersensor tidak memberikan pengaruh yang cukup signifikan terhadap perubahan nilai c-index pada Cox PHM yang diaplikasikan pada data berpola non-linier. Hasil yang sama juga terjadi pada data yang memiliki *sample size* lebih kecil (100) dan data yang memiliki *sample size* lebih besar (5000) dimana nilai c-index dari SURLS-SVM lebih besar dari Cox PHM untuk semua

persentase tersensor, selain itu selisih c-index antar dua model juga lebih signifikan saat diaplikasikan pada data yang berpola non-linier, hal ini juga berlaku untuk jumlah prediktor yang lebih kecil (6,8, dan 15), seperti yang ditampilkan pada Lampiran 4 sampai dengan Lampiran 9 poin (a).

Berdasarkan nilai HR, data berpola linier (Gambar 4.1b), SURLS-SVM memiliki nilai HR yang lebih tinggi dari Cox PHM hanya pada saat persentase tersensor rendah. Sedangkan pada data berpola non-linier (Gambar 4.2b), HR dari SURLS-SVM selalu lebih tinggi dari Cox PHM untuk semua persentase tersensor. Persamaan nilai HR pada data berpola linier dan non-linier adalah keduanya memiliki nilai yang semakin turun saat persentase tersensor naik. Nilai HR pada Cox PHM untuk data non-linier cenderung tetap di semua persentase tersensor dan berada di nilai yang rendah. Nilai HR untuk variasi *sample size* dan jumlah prediktor yang lain dapat dilihat pada Lampiran 4 sampai dengan Lampiran 9 poin (b). Secara umum, pola yang terjadi sama seperti pada *sample size* 1000 dengan jumlah prediktor 17 baik untuk data yang berpola linier maupun data yang berpola non-linier.

Perbandingan performansi yang lain adalah hasil uji log-rank, dimana dalam dunia medis uji log-rank ini menarik untuk memisahkan pasien yang masuk dalam grup *high risk* dan *low risk*. Pada Gambar 4.1(c), dapat dilihat bahwa nilai log-rank dari SURLS-SVM cenderung meningkat saat persentase tersensor juga meningkat. Hal yang sama juga terjadi pada nilai log-rank SURLS-SVM saat model ini diterapkan pada data non-linier. Perbedaan pola kedua jenis data ini adalah, nilai log-rank SURLS-SVM pada data linier lebih tinggi dari Cox PHM di hampir semua persentase tersensor, hanya pada persentase tersensor 10% dan 20% saja nilai log-rank Cox PHM lebih unggul dari SURLS-SVM. Sedangkan pada data non-linier, nilai log-rank SURLS-SVM lebih tinggi dari Cox PHM saat persentase tersensor di atas 50%. Nilai log-rank dari Cox PHM cenderung menurun saat persentase tersensor tinggi, dan penurunan tajam terjadi pada nilai log-rank Cox PHM untuk data non-linier. Nilai log-rank untuk variasi *sample size* dan jumlah prediktor yang lain dapat dilihat pada Lampiran 4 sampai dengan Lampiran 9 poin (c). Secara umum, pola yang terjadi sama seperti pada *sample size* 1000 dengan jumlah

prediktor 17 baik untuk data yang berpola linier maupun data yang berpola non-linier.

Menurut Mahjub *et al.*, (2016), semakin tinggi nilai untuk tiga performansi tersebut, maka semakin baik suatu model. Tetapi, berdasarkan hasil performansi untuk data simulasi hanya c-index yang selalu konsisten memberikan hasil model mana yang lebih baik. Dua ukuran performansi lainnya yaitu HR dan nilai log-rank tidak bisa secara bersamaan menunjukkan hasil yang tinggi atau rendah. Dari semua pola grafik yang terbentuk, saat nilai HR tinggi maka nilai log-rank akan cenderung lebih rendah. Hal ini dapat terjadi karena semakin dekat nilai HR dengan nilai 1, maka semakin besar pula kesamaan antar dua kurva *survival* sehingga nilai log-rank akan turun.

Pada beberapa plot grafik, nilai log-rank cenderung menurun tetapi jika dilihat nilainya masih signifikan sehingga yang terpenting dari nilai log-rank bukan besar atau kecilnya tetapi signifikan atau tidaknya. Jika nilai log-rank sudah signifikan maka model tersebut mampu memisahkan antara grup pasien *low risk* dan pasien *high risk*. Hal lain yang menyebabkan nilai HR dan log-rank tidak lebih stabil dari c-index adalah pendekatan yang digunakan untuk menghitung nilai HR dan log-rank. Dalam menghitung nilai dua performansi tersebut, *prognostic index* dikelompokkan menjadi *high risk* dan *low risk* terlebih dahulu dengan menggunakan batas mean atau median. Hal inilah yang menyebabkan nilai HR dan log-rank kurang stabil karena data hanya diberi label 0 dan 1 tanpa mempertimbangkan seberapa dekat data tersebut dengan nilai mean atau mediannya. Pada c-index, nilai yang digunakan adalah nilai *prognostic index* sebenarnya sehingga mampu menghasilkan pola grafik yang lebih stabil pada berbagai *sample size* dan jumlah prediktor. Jika yang menjadi perhatian adalah prediksi ranking dari *survival time* maka SURLS-SVM akan memberikan hasil yang lebih baik karena nilai c-index yang lebih tinggi dari Cox PHM, artinya SURLS-SVM lebih sedikit mengalami miss-ranking. Tetapi, jika yang menjadi fokus perhatian adalah model yang mampu memisahkan *high risk* dan *low risk*, maka SURLS-SVM tidak selalu memberikan hasil yang diharapkan karena nilai log-rank yang dihasilkan kecil misalkan saat digunakan pada data *sample size* 5000 dengan jumlah prediktor 6 (Lampiran 6.a.), nilai log-rank yang dihasilkan sangat

kecil dan secara statistik tidak memberikan hasil yang signifikan berbeda dan nilai HR pada data tersebut (Lampiran 6.a.) dimana pada setiap persentase tersensor bernilai sekitar 1, sehingga tidak dapat membedakan grup *high risk* dan *low risk*.

Berdasarkan uraian di atas didapatkan suatu informasi bahwa untuk membandingkan dua model dengan pendekatan berbeda dimana dalam penelitian ini membandingkan performansi model dengan pendekatan semi-parametrik (Cox PHM) dan model dengan pendekatan non-parametrik (SURLS-SVM), ukuran yang paling sesuai adalah c-index karena seperti uraian di atas terlihat bahwa c-index memberikan hasil yang konsisten, selain itu ukuran performansi ini mampu menghitung persentase keterurutan antara *survival time* sebenarnya dan *prognostic index* hasil prediksi. Dua ukuran performansi yang lain dapat digunakan untuk melihat performansi model masing-masing saat digunakan untuk memisahkan dua kategori (misalkan *high risk* dan *low risk*), tetapi kurang sesuai jika digunakan untuk membandingkan model karena tolok ukurnya bukan seberapa besar nilainya tetapi apakah nilai itu sudah signifikan atau belum. Lain halnya dengan c-index yang tolok ukurnya adalah nilai, sehingga makin besar nilainya makin baik karena artinya semakin sedikit *prognostic index* yang mis-ranking.

Nilai c-index dari SURLS-SVM pada semua kombinasi jumlah variabel prediktor maupun *sample size* selalu lebih tinggi dari Cox PHM, terutama saat diterapkan pada data yang berpola linier. Hal ini menunjukkan bahwa model SURLS-SVM akan memberikan hasil yang bagus saat diterapkan untuk data non-linier, walaupun pada data linier model ini tetap lebih unggul dari model Cox PHM. Nilai c-index dari SURLS-SVM selalu lebih tinggi dari Cox PHM karena pada saat menentukan *prognostic index* harus melalui pemilihan parameter yang akan memberikan nilai optimal pada c-index sehingga ranking dari *survival time* sangat diperhatikan dan prediksi *prognostic index* menjadi lebih akurat. Hal ini tidak terjadi pada model Cox PHM, dimana pada model ini tidak ada parameter yang dioptimasi. Prediksi *prognostic index* pada model Cox PHM dihasilkan dari perkalian antara penaksir parameter dengan nilai prediktornya sebagaimana dituliskan dalam Bab 2. Padahal penaksir parameter (β) juga memiliki error karena merupakan suatu nilai taksiran, sehingga ketika digunakan untuk menentukan nilai *prognostic index* menjadi kurang sesuai. Akibatnya ketika dihitung menggunakan

c-index ada banyak nilai yang tidak sesuai antara ranking *survival time* dengan ranking *prognostic index* atau terjadi mis-ranking.

Berdasarkan simulasi yang dilakukan dengan menggunakan berbagai skenario, faktor yang paling mempengaruhi perubahan c-index adalah persentase data yang tersensor dimana pada sebagian besar skenario nilai c-index terbesar diperoleh pada saat persentase tersensornya tinggi. *Sample size* dan banyaknya prediktor tidak terlalu memberikan pengaruh yang besar terhadap perubahan nilai c-index, hal ini dapat dilihat dari plot c-index yang hampir mirip untuk *sample size* dan jumlah prediktor yang berbeda-beda,

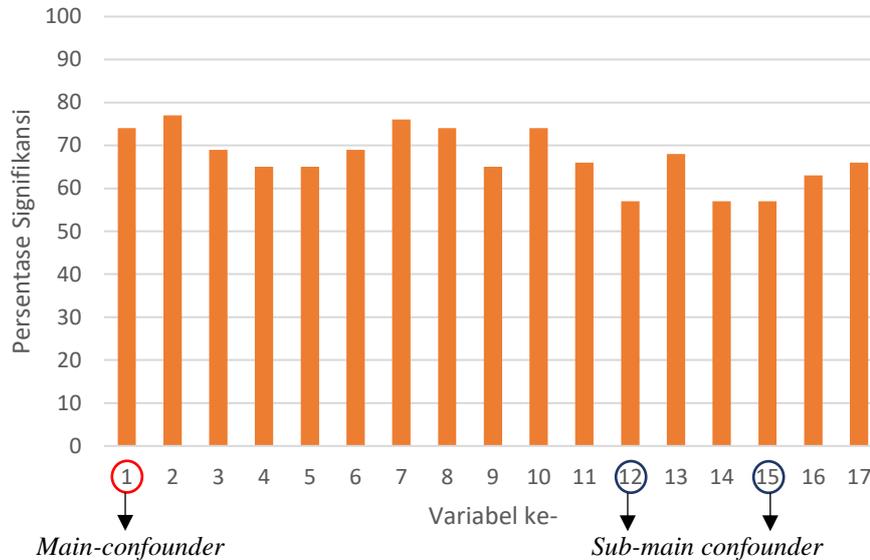
4.2.2 Penerapan *Feature Selection* pada Data Simulasi

Pada skenario membangkitkan data simulasi terdapat satu simulasi yaitu membangkitkan *sample size* 1000 dan 17 variabel prediktor. Dari 17 variabel prediktor terdapat dua variabel (X_{16} dan X_{17}) yang memiliki koefisien sama dengan nol. Hasil yang diharapkan adalah dua variabel tersebut merupakan variabel yang tidak signifikan, dimana dalam penelitian ini dapat dilihat berdasarkan kenaikan c-index saat dua variabel ini tidak masuk dalam model.

Untuk melihat apakah dua variabel tersebut tidak signifikan dapat dilakukan dengan metode eliminasi variabel (*feature selection*) dimana dalam penelitian ini menggunakan *backward selection*. Masing-masing variabel dikeluarkan satu per satu hingga diperoleh c-index terbaik. Nilai terbaik dari c-index diperoleh saat eliminasi variabel tidak menambah nilai c-index tetapi justru membuat nilai c-index lebih kecil. Hasil dari penerapan *feature selection* pada data simulasi lebih akurat jika menggunakan replikasi, karena konsistensi dari dua variabel yang seharusnya tidak signifikan lebih terlihat.

Pada penelitian ini dilakukan simulasi sebanyak 100 kali dengan menggunakan nilai koefisien yang sama pada setiap replikasi untuk *sample size* 1000 dan variabel prediktor 17. Nilai *tuning* parameter yang digunakan pada setiap replikasi adalah sama dengan nilai parameter untuk *full model* (saat semua variabel masuk dalam model). Dalam replikasi *feature selection*, suatu variabel dikatakan signifikan jika variabel tersebut tetap berada dalam model sampai didapatkan nilai c-index paling baik. Dalam proses *feature selection* suatu variabel yang signifikan

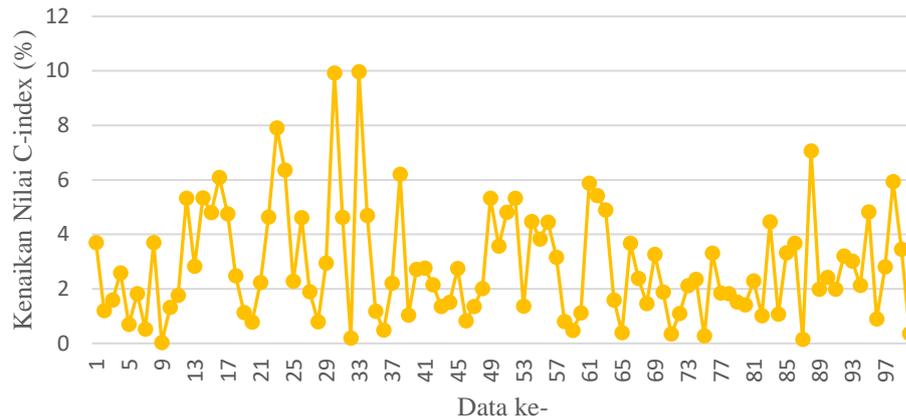
jika dihapus dari model akan menyebabkan nilai c-index berkurang. Berikut ini adalah persentase signifikansi untuk masing-masing variabel.



Gambar 4.3 Persentase Signifikansi Variabel

Berdasarkan hasil pada Gambar 4.3, variabel yang memiliki persentase signifikansi paling tinggi adalah X_2 artinya dalam replikasi sebanyak 100 kali variabel X_2 berada dalam model sebanyak 77%. Variabel X_{16} dan X_{17} merupakan variabel yang memiliki koefisien sama dengan nol. Berdasarkan nilai koefisiennya, seharusnya dua variabel tersebut memiliki persentase signifikansi paling kecil. Pada Gambar 4.3, variabel yang memiliki persentase signifikansi paling rendah adalah X_{12} , sedangkan variabel X_{16} dan X_{17} berturut-turut merupakan variabel terendah keempat dan kedelapan, artinya X_{16} dan X_{17} bukan variabel yang paling banyak keluar dari model.

Selain untuk melihat bagaimana *feature selection* mampu mengidentifikasi variabel yang tidak signifikan, penerapan *feature selection* ini juga dapat digunakan untuk melihat kenaikan nilai c-index setelah metode *backward selection* diterapkan. Berikut ini adalah kenaikan c-index untuk masing-masing replikasi.



Gambar 4.4 Kenaikan Nilai c-index Setelah *Feature Selection*

Pada Gambar 4.4 tampak bahwa pada semua data simulasi kenaikan nilai c-index tidak lebih dari 10%. Banyaknya *feature* yang ada dalam model akhir turut memberikan pengaruh kenaikan nilai c-index. Misalkan dari data simulasi 9 yang memiliki nilai kenaikan c-index paling kecil, *feature selection* pada data tersebut hanya mengeliminasi satu *feature*, sedangkan pada nilai kenaikan c-index terbesar, *feature selection* mengeliminasi enam variabel. Selain itu, pada data yang memiliki nilai kenaikan c-index kecil, X_{16} dan X_{17} tetap berada dalam model sehingga mempengaruhi nilai akhir dari c-index.

Berdasarkan grafik yang ditampilkan pada Gambar 4.3 dan Gambar 4.4 diperoleh hasil bahwa model SURLS-SVM tidak dapat mendeteksi variabel yang sebenarnya tidak signifikan yaitu X_{16} dan X_{17} . Hal ini karena adanya interaksi antar variabel yang juga mempengaruhi persentase signifikansi. Berdasarkan skenario yang ada dalam Bab 3 untuk membangkitkan *survival time* berpola non-linier digunakan variabel yang tidak linier pula karena berinteraksi satu sama lain, yaitu variabel X_{18} yang merupakan hasil interaksi antara $X_1 * X_{15}$ dan X_{19} yang merupakan interaksi antara $X_1 * X_{12}$. Pada Gambar 4.3, variabel yang memiliki persentase signifikansi tinggi salah satunya adalah X_1 yaitu 74%. Jika dilihat berdasarkan interaksinya, variabel X_1 merupakan *main confounder* yang berinteraksi dengan lebih dari satu variabel yaitu dengan variabel X_{12} dan X_{15} sehingga variabel ini sangat mempengaruhi model dan jika X_1 keluar menyebabkan nilai c-index turun. Variabel yang berinteraksi dengan X_1 disebut dengan *sub-main confounder*, variabel ini memiliki probabilitas yang besar untuk dikeluarkan dari

model. Hal ini sesuai dengan hasil yang diperoleh dari replikasi. Variabel X_{12} dan X_{15} memiliki tingkat signifikansi paling rendah diantara variabel yang lain.

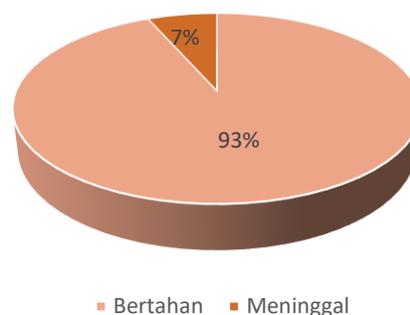
Persentase signifikansi masing-masing variabel memiliki perbedaan yang tidak terlalu jauh, walaupun demikian dari replikasi yang dilakukan efek *feature selection* pada masing-masing data replikasi memberikan hasil yang bervariasi. Dan untuk melihat pengaruh langsung dari variabel yang berinteraksi, maka variabel interaksi tersebut (X_{18} dan X_{19}) perlu dimasukkan sebagai *feature* dalam model SURLS-SVM sehingga akan terlihat bagaimana efek dari interaksi dan efek dari skenario variabel yang memiliki koefisien bernilai nol.

4.3 Analisis Data Pasien Kanker Serviks

Analisis data pada penelitian ini dimulai dengan penggambaran karakteristik pasien dan faktor-faktor yang dianggap berpengaruh terhadap *survival time* pasien kanker serviks, kurva Kaplan-Meier dan uji Log Rank, uji asumsi *Proportional Hazard*, serta analisis dengan Cox PHM dan SURLS-SVM.

4.3.1 Karakteristik pasien kanker serviks

Pasien yang digunakan dalam penelitian ini adalah pasien kanker serviks yang menjalani rawat inap di RSUD dr. Soetomo Surabaya baik yang meninggal maupun yang bertahan hidup. Pasien yang bertahan hidup tersensor karena hingga penelitian berakhir tidak diketahui *survival time* secara pasti. Banyaknya pasien yang bertahan hidup dan meninggal dunia digambarkan pada Gambar 4.5 berikut:



Gambar 4.5 Persentase pasien kanker serviks berdasarkan status tersensor

Berdasarkan Gambar 4.5, pasien yang bertahan 93% atau 385 pasien sedangkan pasien yang meninggal adalah 7% atau 27 pasien.

Berikut ini karakteristik pasien kanker serviks berdasarkan faktor-faktor yang diduga mempengaruhi waktu bertahan pasien kanker serviks:

1. Karakteristik pasien kanker serviks berdasarkan usia pasien, usia saat pertama menikah, dan usia saat menstruasi pertama.

Statistika deskriptif untuk masing-masing faktor disajikan pada Tabel 4.2.

Tabel 4.2 Statistika deskriptif faktor usia, usia saat pertama menikah, dan usia saat menstruasi pertama

Deskriptif	Usia (tahun)		Usia pertama menikah (tahun)		Usia saat menstruasi pertama (tahun)	
	Bertahan	Meninggal	Bertahan	Meninggal	Bertahan	Meninggal
Rata-rata	49,15	50,04	18,93	20,11	13,29	13,11
Maksimum	80	69	34	30	18	17
Median	49	51	19	19	13	13
Minimum	29	33	0	15	9	11
St.deviasi	8,38	8,12	3,74	3,76	1,51	1,31

Tabel 4.2 menunjukkan bahwa rata-rata usia pasien yang terkena kanker serviks adalah sekitar 49 tahun, dengan usia maksimum untuk pasien yang bertahan hidup dan meninggal berturut-turut adalah 80 dan 69 tahun, usia minimum pasien yang terkena kanker serviks adalah 29 tahun untuk pasien yang bertahan hidup dan 33 tahun untuk pasien yang meninggal, dan standard deviasi usia untuk pasien yang bertahan hidup dan meninggal adalah sekitar 8. Berdasarkan faktor usia saat pertama kali menikah, pasien yang meninggal memiliki rata-rata usia saat pertama kali menikah yang lebih tinggi dibandingkan pasien yang bertahan hidup, nilai maksimum usia saat pertama menikah untuk pasien yang bertahan adalah 34 tahun dan untuk pasien yang meninggal adalah 30 tahun, nilai minimum usia saat pertama menikah untuk pasien yang bertahan adalah 0 tahun artinya pasien tersebut tidak menikah dan nilai minimum untuk pasien yang meninggal adalah 15 tahun, standard deviasi untuk pasien yang bertahan hidup dan pasien yang meninggal adalah sekitar 3,7. Berdasarkan faktor usia saat menstruasi pertama, rata-rata untuk pasien yang bertahan hidup dan meninggal adalah sekitar 13 tahun, nilai-nilai statistika deskriptif lainnya untuk pasien yang meninggal dan bertahan hidup hampir sama.

2. Karakteristik pasien kanker serviks berdasarkan siklus menstruasi, lama menstruasi, dan paritas.

Statistika deskriptif untuk masing-masing faktor disajikan secara lengkap pada Tabel 4.3.

Tabel 4.3. Statistika deskriptif faktor siklus menstruasi, lama menstruasi, dan paritas

Deskriptif	Siklus menstruasi (hari)		Lama menstruasi (hari)		Paritas	
	Bertahan	Meninggal	Bertahan	Meninggal	Bertahan	Meninggal
Rata-rata	28,42	28,67	6,49	6,22	2,72	2,81
Maksimum	35	30	12	7	8	6
Median	28	28	7	7	3	3
Minimum	18	28	3	3	0	0
St.deviasi	1,24	0,96	1,17	1,34	1,45	1,49

Tabel 4.3 menunjukkan rata-rata siklus menstruasi untuk pasien yang bertahan hidup maupun yang meninggal adalah sekitar 28 hari. Berdasarkan faktor lama menstruasi untuk satu periode, rata-rata untuk pasien yang bertahan hidup dan meninggal adalah 6 hari, dengan nilai maksimum untuk pasien yang bertahan adalah 12 hari dan untuk pasien yang meninggal adalah 7 hari, nilai minimum dan standard deviasi untuk pasien yang bertahan hidup maupun yang meninggal adalah sama, berturut-turut yaitu 3 hari dan 1. Berdasarkan faktor paritas, rata-rata pasien yang bertahan dan meninggal memiliki 3 orang anak, nilai maksimum untuk pasien yang bertahan hidup memiliki 8 orang anak dan pasien yang meninggal memiliki 6 orang anak, pasien yang bertahan hidup dan pasien yang meninggal ada yang tidak memiliki anak sehingga nilai minimumnya adalah nol, dan standard deviasi paritas untuk pasien yang bertahan hidup maupun pasien yang meninggal adalah sekitar 1,4.

3. Karakteristik pasien kanker serviks berdasarkan faktor komplikasi

Salah satu faktor yang diduga mempengaruhi kelangsungan hidup (*survival time*) pasien kanker serviks adalah komplikasi. Karakteristik pasien kanker serviks yang menjalani rawat inap di RSUD dr. Soetomo berdasarkan faktor komplikasi dijelaskan pada Tabel 4.4.

Tabel 4.4. Tabulasi silang antara status pasien dengan status komplikasi

	Ada	Tidak ada	Total
Bertahan	292	93	385
	70,87%	22,58%	93,45%
<i>Expected Value</i>	294,4	90,6	385
Meninggal	23	4	27
	5,58%	0,97%	6,55%
<i>Expected Value</i>	20,6	6,4	27
Total	315	97	412
	76,46%	23,54%	100%

Tabel 4.4 memberikan informasi bahwa baik pasien yang bertahan maupun yang meninggal kebanyakan mengalami komplikasi dengan penyakit lain selain kanker serviks yaitu berturut-turut 70,87% dan 5,58%. Total pasien yang mengalami komplikasi penyakit lain adalah 76,46% atau 315 pasien, sedangkan pasien yang tidak mengalami komplikasi adalah 97 pasien atau sekitar 23,54%. Persentase paling kecil terjadi pada pasien yang meninggal dan tidak mengalami komplikasi penyakit lain selain kanker serviks.

4. Karakteristik pasien kanker serviks berdasarkan faktor anemia

Anemia atau penyakit kekurangan sel darah merah menjadi salah satu penyakit penyerta dari pasien kanker serviks tetapi tidak semua pasien kanker serviks mengalami anemia selama menjalani pengobatan dan rawat inap. Karakteristik pasien kanker serviks dilihat berdasarkan status anemia dijelaskan pada Tabel 4.5.

Tabel 4.5. Tabulasi silang antara status pasien dengan status anemia

	Ya	Tidak	Total
Bertahan	242	143	385
	58,74%	34,71%	93,45%
<i>Expected Value</i>	247,6	137,4	385
Meninggal	23	4	27
	5,58%	0,97%	6,55%
<i>Expected Value</i>	17,4	9,6	27
Total	265	147	412
	64,32%	35,68%	100%

Berdasarkan Tabel 4.5, pasien yang mengalami anemia adalah 265 pasien atau sekitar 64,32%, dan pasien yang tidak pernah mengalami anemia selama menjalani rawat inap adalah 147 pasien atau sekitar 35,68%. Pasien yang meninggal dan

mengalami anemia sebanyak 23 pasien atau sekitar 85,19% dari total pasien yang meninggal. Persentase pasien yang paling sedikit terjadi pada pasien yang meninggal tetapi tidak mengalami anemia. Sebagian besar pasien kanker serviks yang bertahan hidup juga pernah mengalami anemia selama menjalani rawat inap.

5. Karakteristik pasien kanker serviks berdasarkan jenis pengobatan

Jenis pengobatan juga diduga menjadi faktor kelangsungan hidup dari pasien kanker serviks. Berdasarkan data rekam medis pasien kanker serviks yang menjalani rawat inap, jenis pengobatan yang diberikan adalah kemoterapi, transfuse PRC, dan operasi. Karakteristik pasien kanker serviks jika dilihat berdasarkan faktor jenis pengobatan yang pernah dijalani dijelaskan pada Tabel 4.6.

Tabel 4.6. Tabulasi silang antara status pasien dengan jenis pengobatan

	Kemoterapi	Transfusi PRC	Kemoterapi dan Transfusi PRC	Lainnya	Total
Bertahan	181 43,93%	29 7,04%	153 37,14%	22 5,34%	385 93,45%
<i>Expected Value</i>	175,7	32,7	152,3	24,3	385
Meninggal	7 1,69%	6 1,46%	10 2,43%	4 0,97%	27 6,55%
<i>Expected Value</i>	12,3	2,3	10,7	1,7	27
Total	188 45,63%	35 8,49%	163 39,56%	26 6,31%	412 100%

Berdasarkan Tabel 4.6 secara keseluruhan jenis pengobatan yang paling sering diberikan kepada pasien kanker serviks adalah kemoterapi sebesar 45.63% dari total semua pasien, dan yang paling sedikit adalah operasi yaitu hanya sekitar 6.31% dari total semua pasien. Pasien yang meninggal paling banyak mendapat jenis pengobatan transfusi PRC yang dilanjutkan dengan kemoterapi yaitu sekitar 37.04% dari total semua pasien yang meninggal. Pasien yang bertahan hidup kebanyakan menjalani jenis pengobatan kemoterapi dengan persentase sekitar 47% dari total pasien yang bertahan hidup.

6. Karakteristik pasien kanker serviks berdasarkan faktor stadium

Stadium merupakan faktor yang mempengaruhi kelangsungan hidup pasien kanker serviks karena stadium merupakan salah satu tolok ukur keparahan dari penyakit kanker. Pada Lampiran 3.a ditampilkan tabulasi silang untuk data stadium sebenarnya dari pasien kanker serviks, tetapi jika dilihat nilai *expected value*

masing-masing kategori ada satu kategori yang memiliki *expected value* kurang dari 0.5 sehingga berdasarkan syarat pembentukan kategori tidak memenuhi. Selain itu dari kategori yang terbentuk terdapat lebih dari 20% kombinasi kategori memiliki *expected value* kurang dari 5. Untuk mengatasi hal ini maka ada kategori yang digabung menjadi satu agar syarat penentuan kategori terpenuhi. Dalam penelitian ini, stadium 1 digabung dengan stadium 2 (diberi kode 0) dan stadium 3 digabung dengan stadium 4 (diberi kode 1) karena hanya sedikit pasien yang memiliki stadium 1 dan stadium 4. Karakteristik pasien kanker serviks baik yang bertahan hidup maupun yang meninggal berdasarkan stadium yang baru dijelaskan pada Tabel 4.7.

Tabel 4.7. Tabulasi silang antara status pasien dengan stadium

	0 (Stadium 1 dan 2)	1 (Stadium 3 dan 4)	Total
Bertahan	189 45,87%	196 45,57%	385 93,44%
<i>Expected Value</i>	180,4	204,6	385
Meninggal	4 0,97%	23 5,59%	27 6,56%
<i>Expected Value</i>	12,6	14,4	27
Total	193 46,84%	219 53,16%	412 100%

Pada Tabel 4.7 tampak bahwa pasien kanker serviks paling banyak adalah stadium 3 ke atas. Pasien meninggal yang lebih banyak juga terjadi pada stadium 3 ke atas.

7. Karakteristik pasien kanker serviks berdasarkan faktor jenis KB

Jenis KB yang dilakukan oleh pasien kanker serviks adalah hormonal (suntik dan pil) serta non-hormonal (pemasangan IUD), tetapi ada juga beberapa pasien yang tidak melakukan KB. Karakteristik pasien kanker serviks berdasarkan status pasien dan jenis KB yang dilakukan disajikan pada Tabel 4.8.

Tabel 4.8. Tabulasi silang antara status pasien dengan jenis KB

	Tidak KB	KB Suntik	KB Pil	IUD	Total
Bertahan	146 35,44%	114 27,67%	67 16,26%	58 14,08%	385 93,45%
<i>Expected Value</i>	149,5	114	64,5	57	385
Meninggal	14 3,39%	8 1,94%	2 0,49%	3 0,73%	27 6,55%
<i>Expected Value</i>	10,5	8	4,5	4	27
Total	160 38,83%	122 29,61%	69 16,75%	61 14,81%	412 100%

Berdasarkan Tabel 4.8, persentase pasien yang tidak KB adalah paling besar baik untuk pasien yang bertahan hidup maupun pasien yang meninggal yaitu 160 pasien atau 38.83% dari total semua pasien. Dari tiga jenis KB yang paling banyak digunakan oleh pasien adalah KB suntik yaitu sebesar 48.41% dibandingkan dengan jenis KB yang lain baik untuk pasien yang bertahan hidup maupun pasien yang meninggal. KB yang paling sedikit diambil oleh pasien yang bertahan hidup adalah IUD yaitu 58 pasien dan untuk pasien yang meninggal paling sedikit memilih untuk melakukan KB suntik yaitu 2 pasien.

8. Karakteristik pasien kanker serviks berdasarkan faktor pendidikan

Pendidikan diduga menjadi salah satu faktor yang juga menentukan bagaimana kelangsungan hidup kanker serviks, karena seharusnya semakin tinggi pendidikan seharusnya semakin mengetahui mengenai pengobatan kanker serviks. Pada Lampiran 3.b. disajikan kategori tingkat pendidikan berdasarkan data sebenarnya. Salah satu kategori yaitu pasien dengan tingkat pendidikan perguruan tinggi dan status meninggal mempunyai *expected value* kurang dari 1 sehingga tidak memenuhi kriteria untuk membentuk kategori. Oleh karena itu, pasien dengan pendidikan SMA dan perguruan tinggi digabung dan diberi kode 2. Karakteristik pasien kanker serviks baik yang bertahan hidup maupun yang meninggal berdasarkan faktor pendidikan yang baru ditampilkan pada Tabel 4.9.

Tabel 4.9. Tabulasi silang antara status pasien dengan tingkat pendidikan

	0 SD	1 SMP	2 SMA dan Perguruan Tinggi	Total
Bertahan	121 29,4%	60 14,6%	204 49,5%	385 93,5%
<i>Expected Value</i>	117,7	58,9	208,4	385
Meninggal	5 1,2%	3 0,7%	19 4,6%	27 6,5%
<i>Expected value</i>	8,3	4,1	14,6	27
Total	126 30,58%	63 15,29%	217 52,67%	412 100%

Berdasarkan Tabel 4.9 diketahui bahwa tingkat pendidikan pasien kanker serviks yang paling banyak adalah SMA ke atas yaitu 217 pasien atau sekitar 52,67% dan yang paling banyak meninggal juga yang memiliki pendidikan SMA ke atas.

Untuk melihat hubungan antara status pasien dengan masing-masing variabel prediktor kategorik digunakan uji asosiasi dengan menggunakan uji statistik *chi-square*. Hasil dari uji asosiasi ditampilkan pada Tabel 4.10 berikut:

Tabel 4.10 Uji Asosiasi Variabel

Variabel	df	χ^2	<i>p-value</i>
Komplikasi	1	1,22	0,27
Anemia	1	5,48	0,02
Jenis Pengobatan	3	12,23	0,01
Stadium	1	11,90	0,00
Jenis KB	3	3,03	0,39
Tingkat Pendidikan	2	3,11	0,21

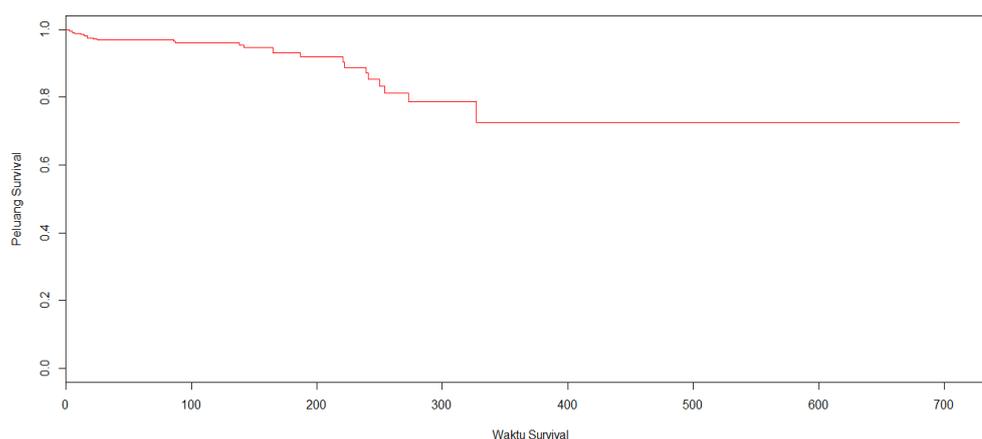
Berdasarkan Tabel 4.10 terdapat tiga variabel yang menolak H_0 yaitu anemia, jenis pengobatan, dan stadium, artinya terdapat hubungan signifikan antara status pasien (bertahan atau meninggal) dan salah satu variabel tersebut.

4.3.2 Analisis kurva *survival* Kaplan Meier dan uji Log-Rank

Kurva *survival* Kaplan Meier digunakan untuk melihat bagaimana karakteristik *survival* dari pasien kanker serviks berdasarkan faktor-faktor yang diduga mempengaruhi *survival time* pasien, yaitu usia, komplikasi, anemia, jenis pengobatan, stadium, umur saat menikah pertama, umur saat menstruasi pertama, siklus menstruasi, lama menstruasi, paritas, jenis KB, dan tingkat pendidikan. Uji

log-rank digunakan untuk melihat apakah ada perbedaan kurva *survival* dalam satu faktor yang terbagi menjadi beberapa kategori. Untuk faktor usia, umur saat menikah pertama, umur saat menstruasi pertama, siklus menstruasi, lama menstruasi, dan paritas merupakan data kontinu sehingga sebelum dibuat kurva Kaplan Meier dan uji log-rank diubah ke dalam nilai diskrit berdasarkan nilai rata-rata pada masing-masing variabel.

Pola dari kurva *survival* pada pasien kanker serviks tanpa melibatkan faktor-faktor yang mempengaruhi ditampilkan pada Gambar 4.6.

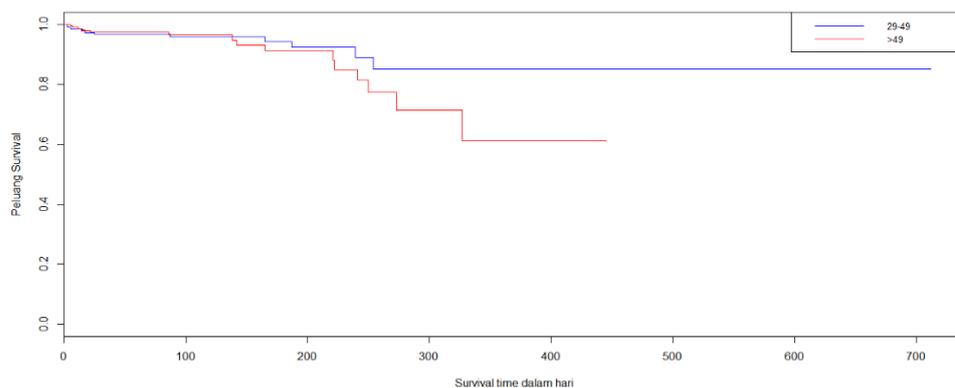


Gambar 4.6. Kurva Kaplan Meier pasien kanker serviks

Pada Gambar 4.6 dapat dilihat bahwa pada hari ke-0 sampai dengan hari ke-310, ketahanan hidup pasien rawat inap menurun secara perlahan tetapi masih cukup tinggi (di atas 0.7). Setelah hari ke-310, peluang *survival* cenderung stabil sampai dengan pengamatan selesai. Selanjutnya, gambaran *survival time* berdasarkan faktor-faktor yang diduga mempengaruhi dijelaskan sebagai berikut:

1. Kurva *survival* Kaplan Meier pada faktor usia

Kurva *survival* Kaplan Meier untuk faktor usia disajikan pada Gambar 4.7.

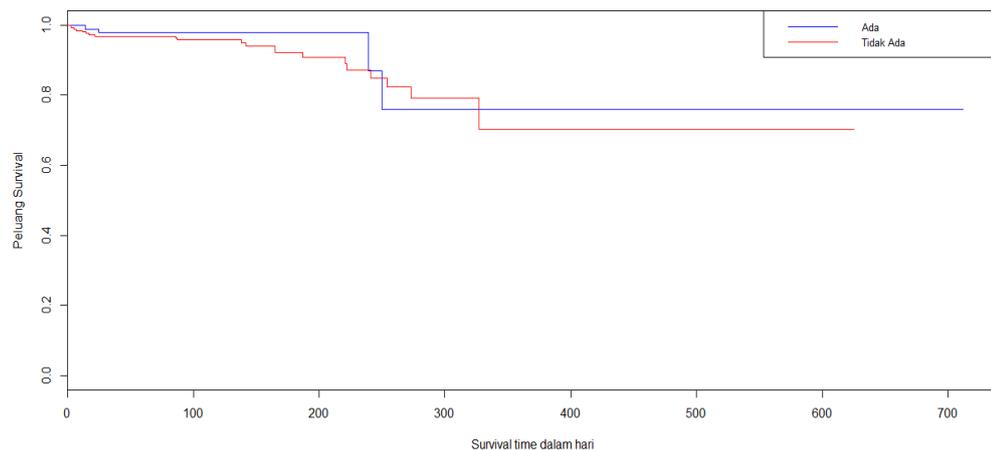


Gambar 4.7 Kurva Kaplan Meier berdasarkan usia

Berdasarkan Gambar 4.7 tampak bahwa pasien dengan usia di bawah atau sama dengan 49 tahun dan pasien berusia di atas 49 tahun pada awal *survival time* dua kelompok usia tersebut berhimpit. Tetapi semakin bertambahnya *survival time*, kelompok usia di atas 49 tahun menurun cukup tajam saat hari ke-210, sedangkan kelompok usia yang lain menurun secara perlahan dan cenderung stabil pada hari ke-210. Dengan kata lain, pasien dengan usia di atas 49 tahun memiliki peluang bertahan lebih rendah dibanding dengan kelompok usia di bawah 49 tahun.

2. Kurva *survival* Kaplan-Meier pada faktor komplikasi

Kurva *survival* Kaplan Meier untuk faktor usia disajikan pada Gambar 4.8.

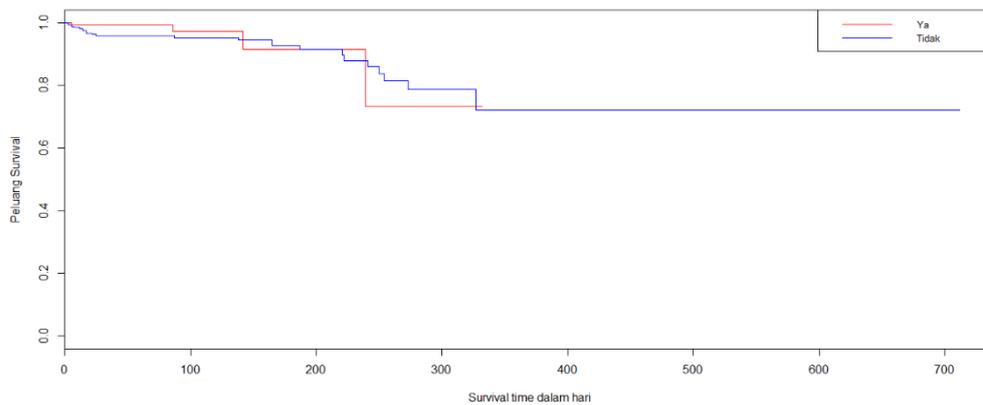


Gambar 4.8 Kurva Kaplan Meier berdasarkan komplikasi

Berdasarkan Gambar 4.8, kurva *survival* pasien yang memiliki komplikasi dengan pasien yang tidak memiliki komplikasi saling berpotongan dan setelah hari ke-300, kurva *survival* pasien yang tidak memiliki komplikasi lebih rendah dibandingkan dengan pasien yang memiliki komplikasi. Namun secara umum, kurva *survival* kedua kelompok relatif sama.

3. Kurva *survival* Kaplan-Meier pada faktor anemia

Kurva *survival* Kaplan-Meier untuk faktor anemia disajikan pada Gambar 4.9.

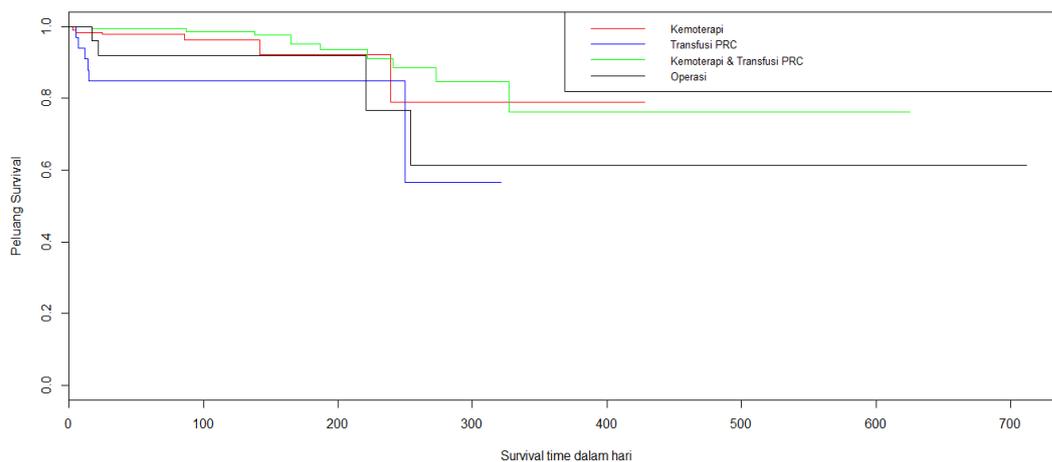


Gambar 4.9. Kurva Kaplan Meier berdasarkan anemia

Berdasarkan Gambar 4.9 dapat dilihat bahwa kurva *survival* antara kelompok pasien yang terkena anemia dan pasien yang tidak terkena anemia saling berpotongan. Setelah hari ke-300, peluang ketahanan hidup dua kelompok tersebut cenderung konstan yaitu sekitar 0.7. Dua kurva *survival* tersebut memiliki pola yang hampir sama walaupun kelompok pasien yang menderita anemia turun secara tajam.

4. Kurva *survival* Kaplan-Meier pada faktor jenis pengobatan

Kurva *survival* Kaplan-Meier untuk faktor jenis pengobatan ditampilkan pada Gambar 4.10.



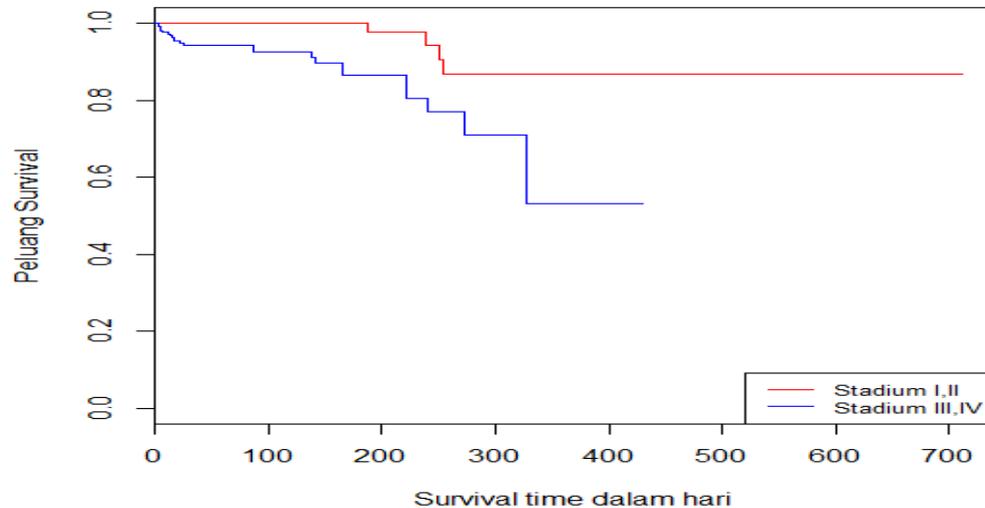
Gambar 4.10. Kurva Kaplan-Meier berdasarkan jenis pengobatan

Berdasarkan Gambar 4.10, masing-masing jenis pengobatan mempunyai kurva *survival* yang berbeda-beda. Berdasarkan peluang bertahan, pengobatan kemoterapi dan transfusi PRC yang dilakukan bersamaan memberikan peluang bertahan paling besar dibandingkan jenis pengobatan lainnya. Berdasarkan lamanya pasien bertahan, jenis pengobatan operasi memiliki *survival time* paling panjang walaupun peluang bertahannya masih di bawah dua

jenis pengobatan yang lain. Jenis pengobatan yang memiliki *survival time* terpendek sekaligus peluang bertahan yang kecil adalah transfusi PRC.

5. Kurva *survival* Kaplan-Meier pada faktor stadium

Kurva *survival* Kaplan-Meier untuk stadium ditampilkan pada Gambar 4.11

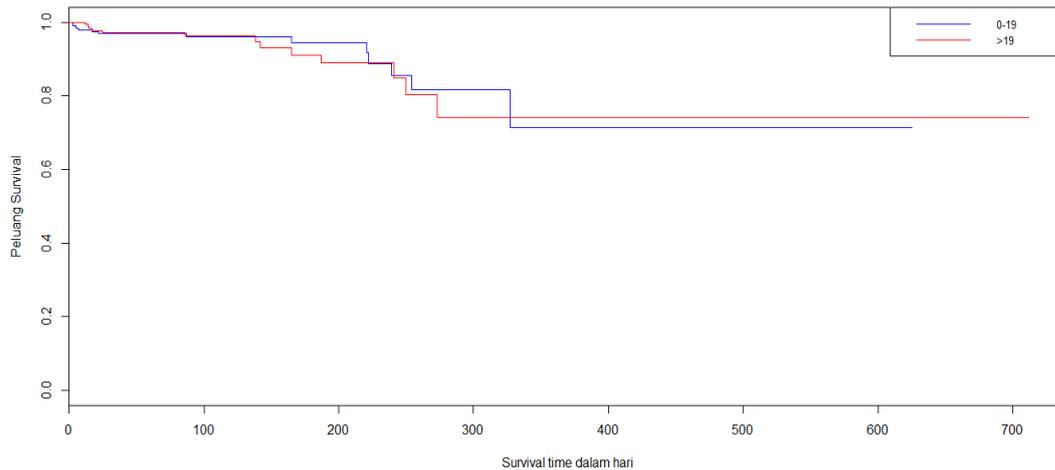


Gambar 4.11. Kurva Kaplan-Meier berdasarkan stadium

Berdasarkan Gambar 4.11, tampak bahwa mulai hari pertama sampai hari ke-700, kurva *survival* untuk gabungan pasien stadium I dan II lebih tinggi dari pasien stadium III, IV. Hal ini sesuai dengan teori bahwa semakin tinggi stadium kanker maka peluang penderitanya untuk bertahan semakin kecil. Pada Gambar 4.11 juga terlihat bahwa kurva *survival* untuk pasien stadium III,IV menurun lebih cepat dan lebih pendek daripada kurva *survival* pasien stadium I,II.

6. Kurva *survival* Kaplan-Meier pada faktor umur saat menikah pertama

Kurva *survival* Kaplan-Meier untuk faktor umur saat menikah pertama ditampilkan pada Gambar 4.12.

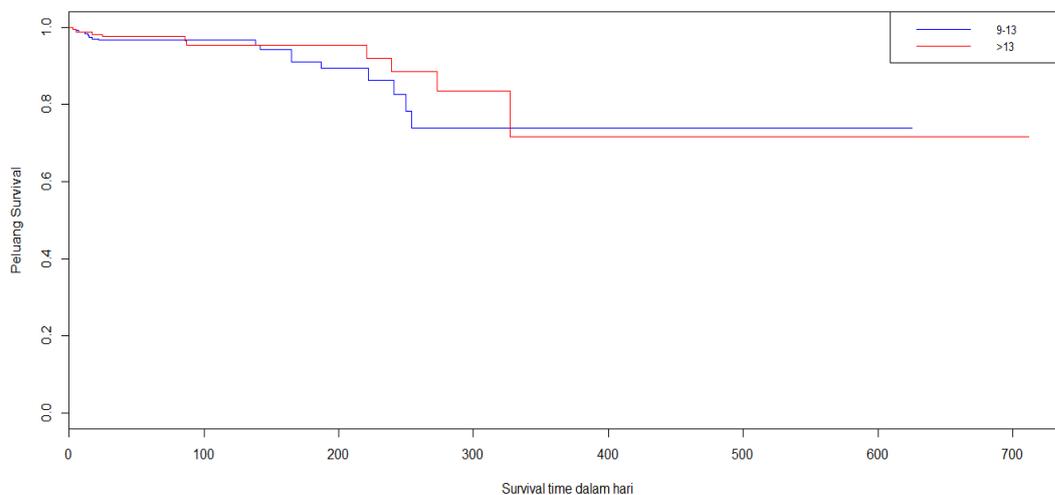


Gambar 4.12 Kurva Kaplan-Meier berdasarkan usia saat menikah pertama

Usia saat pertama menikah merupakan variabel kontinu sehingga dari data pasien kanker serviks dibagi menjadi dua kelompok berdasarkan nilai rata-ratanya. Berdasarkan Gambar 4.12, kelompok yang menikah saat berusia kurang dari atau sama dengan 19 tahun dan kelompok yang menikah di atas 19 tahun memiliki kurva *survival* yang berhimpit dan menurun perlahan. Kelompok yang menikah di atas 19 tahun memiliki *survival time* yang lebih panjang dan peluang bertahan hidup yang lebih besar dibandingkan dengan kelompok lainnya.

7. Kurva *survival* Kaplan-Meier pada faktor umur saat menstruasi pertama

Kurva *survival* Kaplan-Meier untuk faktor umur saat menstruasi pertama ditampilkan pada Gambar 4.13.



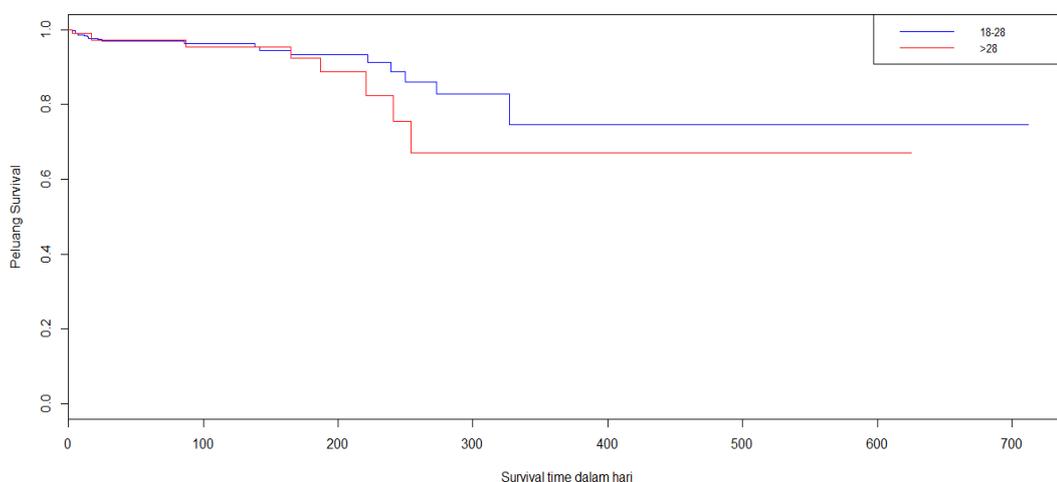
Gambar 4.13. Kurva Kaplan-Meier berdasarkan usia saat menstruasi pertama

Usia saat menstruasi pertama dari pasien kanker serviks dibagi menjadi dua kelompok berdasarkan rata-ratanya. Berdasarkan Gambar 4.13 tampak bahwa

kurva *survival* antara kelompok yang menstruasi pertama di bawah atau sama dengan 19 tahun dengan kelompok di atas 19 tahun saling berhimpit dan menurun secara perlahan. Artinya semakin panjang *survival time*, peluang bertahan hidup kedua kelompok semakin berkurang. Pasien yang mengalami menstruasi pertama di atas 13 tahun memiliki *survival time* yang lebih panjang dibandingkan dengan kelompok lainnya, tetapi peluang bertahannya lebih tinggi pasien yang mengalami menstruasi di bawah atau sama dengan 13 tahun.

8. Kurva *survival* Kaplan-Meier pada faktor siklus menstruasi

Kurva *survival* Kaplan-Meier untuk faktor siklus menstruasi ditampilkan pada Gambar 4.14.

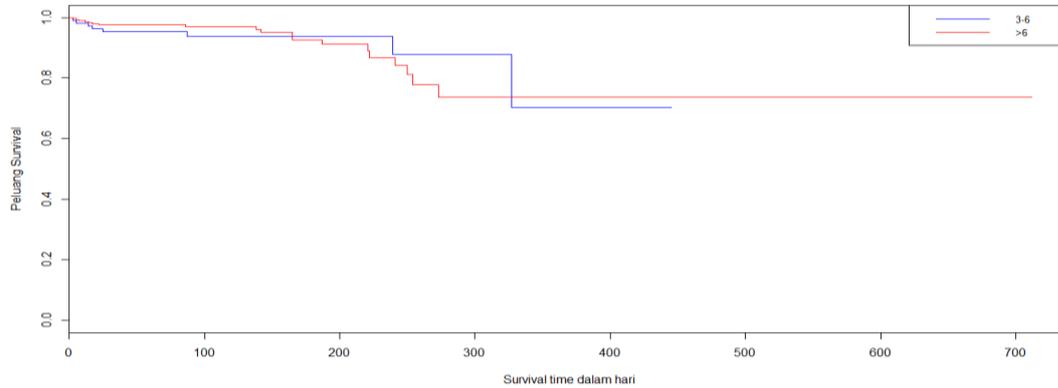


Gambar 4.14 Kurva Kaplan-Meier berdasarkan siklus menstruasi

Siklus menstruasi dibagi menjadi dua kelompok berdasarkan nilai rata-ratanya, yaitu kelompok yang memiliki siklus antara 18 hari sampai dengan 28 hari dan kelompok yang memiliki siklus di atas 28 hari. Berdasarkan Gambar 4.14, kedua kelompok memiliki kurva *survival* yang saling berhimpit dan semakin bertambah *survival time*, peluang bertahan semakin menurun perlahan. Setelah 220 hari kurva *survival* untuk pasien yang memiliki siklus lebih dari 28 hari sudah mulai stabil, sedangkan satu kelompok yang lain mulai stabil setelah hari ke-310. Kelompok pasien yang memiliki siklus antara 18 hari sampai dengan 28 hari memiliki *survival time* yang lebih panjang dan peluang bertahan yang lebih tinggi dibandingkan dengan pasien yang memiliki siklus menstruasi lebih dari 28 hari.

9. Kurva *survival* Kaplan-Meier pada faktor lama menstruasi dalam satu periode

Kurva *survival* Kaplan-Meier untuk faktor lama menstruasi ditampilkan pada Gambar 4.15.

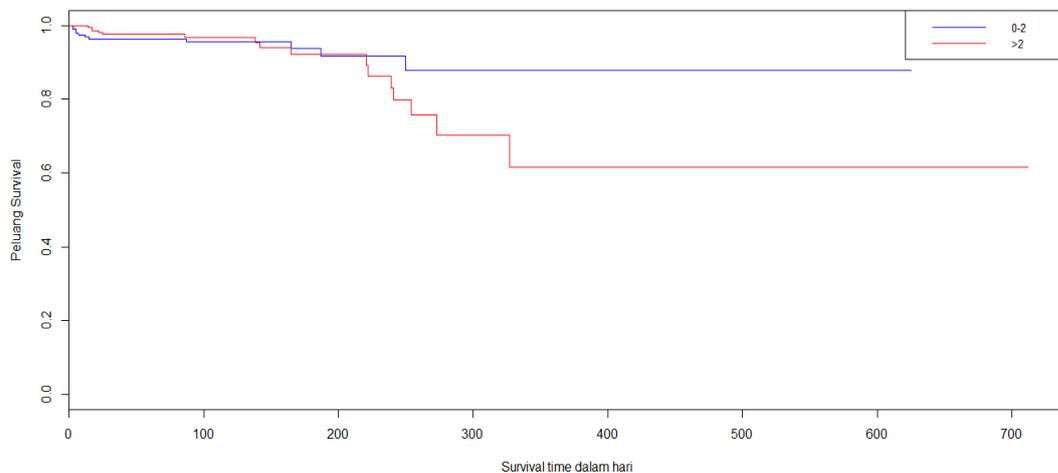


Gambar 4.15 Kurva *survival* Kaplan-Meier berdasarkan lama menstruasi

Lama menstruasi dalam satu periode dibagi menjadi dua kelompok berdasarkan nilai rata-ratanya, yaitu kelompok pasien yang memiliki lama menstruasi dari 3 hari sampai 6 hari dan kelompok yang memiliki menstruasi lebih dari 6 hari. Berdasarkan Gambar 4.15, kurva *survival* untuk kedua kelompok saling berhimpit dan tidak memiliki perbedaan yang signifikan, walaupun pasien yang memiliki lama menstruasi lebih dari 6 hari memiliki *survival time* lebih pendek dan peluang bertahan yang lebih kecil.

10. Kurva *survival* Kaplan-Meier pada faktor paritas

Kurva *survival* Kaplan-Meier untuk faktor paritas ditampilkan pada Gambar 4.16.

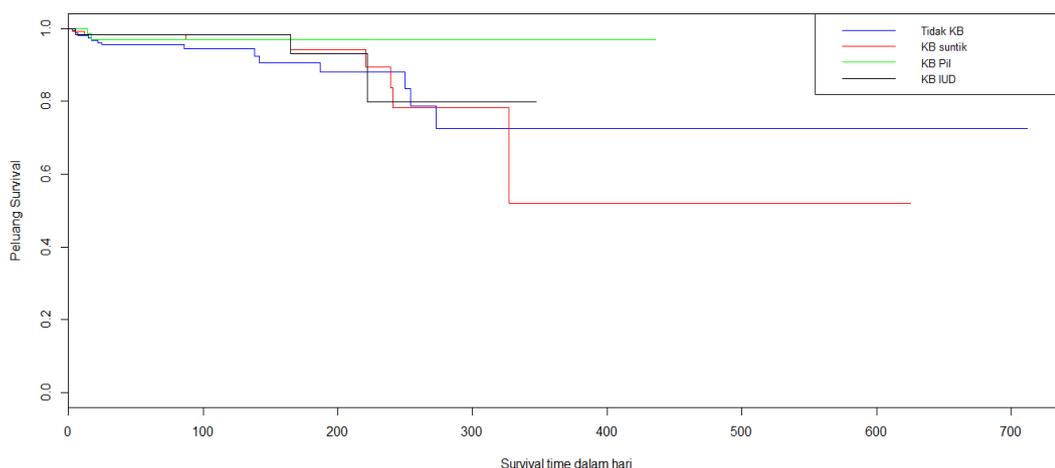


Gambar 4.16 Kurva Kaplan-Meier berdasarkan paritas

Paritas dibagi menjadi dua kelompok berdasarkan rata-ratanya, yaitu kelompok yang memiliki paritas 0,1, dan 2 serta kelompok yang memiliki paritas lebih dari 2. Berdasarkan Gambar 4.16, kurva *survival* dari kedua kelompok saling berhimpit dan hampir sama walaupun kurva untuk pasien dengan paritas lebih dari 2 memiliki *survival time* lebih panjang dan peluang bertahan yang lebih kecil daripada pasien dengan paritas 0,1,2.

11. Kurva *survival* Kaplan-Meier pada faktor jenis KB

Kurva *survival* Kaplan-Meier untuk faktor jenis KB ditampilkan pada Gambar 4.17.

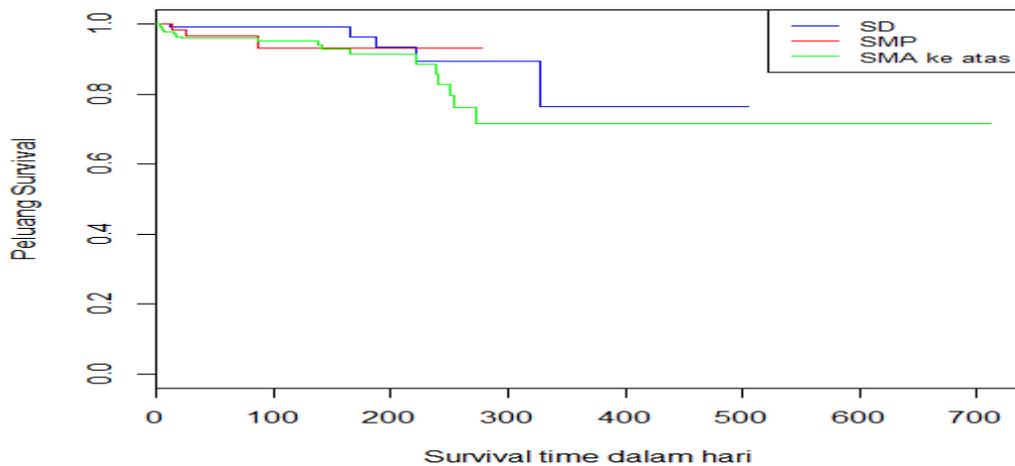


Gambar 4.17. Kurva Kaplan-Meier berdasarkan jenis KB

Berdasarkan Gambar 4.17, kurva *survival* untuk masing-masing kelompok saling berhimpit dan menurun perlahan saat *survival time* bertambah. Pasien yang menjalani KB suntik memiliki kurva *survival* yang menurun cukup tajam setelah hari ke-300, pasien yang menjalani KB Pil memiliki kurva *survival* yang cenderung stabil dan memiliki peluang bertahan hidup paling tinggi, pasien yang tidak KB walaupun peluang bertahan hidup menurun tetapi memiliki *survival time* yang paling panjang, dan pasien yang menjalani KB IUD memiliki *survival time* yang pendek dibandingkan dengan jenis KB yang lain.

12. Kurva *survival* Kaplan-Meier pada faktor tingkat pendidikan

Kurva *survival* Kaplan-Meier berdasarkan faktor pendidikan ditampilkan pada Gambar 4.18.



Gambar 4.18 Kurva Kaplan Meier berdasarkan tingkat pendidikan

Gambar 4.18 menggambarkan bahwa kurva *survival* berdasarkan tingkat pendidikan saling berhimpit dan menurun secara perlahan. Pasien dengan pendidikan SMA ke atas memiliki jumlah terbanyak dan *survival time* terpanjang, serta mulai konstan setelah hari ke-250. Pasien dengan pendidikan SD terbanyak kedua setelah SMA ke atas dan memiliki kurva *survival* yang turun cukup tajam setelah hari ke-300. Pasien dengan pendidikan SMP memiliki kurva *survival* yang mulai konstan setelah hari ke-100 dan memiliki peluang bertahan hidup cukup tinggi dibandingkan dengan tingkat pendidikan lainnya.

Kurva *survival* Kaplan-Meier dapat digunakan untuk melihat secara visual apakah kurva *survival* dari dua kelompok sama atau berbeda. Secara statistik, untuk melihat apakah berbeda atau sama dapat digunakan uji Log-rank. Tabel 4.11 menjelaskan hasil uji log-rank dari dua belas variabel yang digunakan dalam data.

Tabel 4.11 Hasil uji log-rank per variabel

Variabel	Nilai Log-rank	df	p-value
Usia	0,80	1	0,365
Komplikasi	0,50	1	0,490
Anemia	0,90	1	0,340
Jenis pengobatan	14,3	3	0,003
Stadium	14,5	1	0,000
Usia saat menikah pertama	0,20	1	0,677
Usia saat menstruasi pertama	0,70	1	0,450
Siklus menstruasi	1,00	1	0,320
Lama menstruasi dalam satu periode	0,00	1	0,965
Paritas	0,60	1	0,437
Jenis KB	2,40	3	0,498
Tingkat pendidikan	3,10	2	0,210

Berdasarkan Tabel 4.11 variabel yang memiliki kurva *survival* berbeda antar kelompok adalah jenis pengobatan dan stadium. Dengan kata lain pemberian jenis pengobatan yang berbeda menyebabkan perbedaan *survival time* dan peluang bertahan hidup secara signifikan. Tingkatan stadium pasien kanker serviks juga memberikan perbedaan yang signifikan pada *survival time* dan peluang bertahan hidup secara signifikan.

4.4 Analisis Data Pasien Kanker Serviks dengan Cox PHM dan SURLS-SVM

Pada subbab ini dibahas mengenai analisis data pasien kanker serviks menggunakan Cox PHM yang diawali dengan uji proportional hazard pada setiap variabel yang diduga mempengaruhi dan dilanjutkan dengan pembentukan model menggunakan Cox PHM untuk mengetahui variabel yang secara signifikan mempengaruhi *survival time* pasien kanker serviks.

4.4.1 Uji asumsi proporsional hazard pada setiap variabel

Uji asumsi proporsional hazard (PH) dilakukan untuk melihat apakah setiap variabel yang diduga berpengaruh memenuhi asumsi atau tidak. Pengujian asumsi PH dilakukan dengan menggunakan uji statistik yaitu uji korelasi antara Schoenfeld residual dengan ranking dari *survival time*. Hasil dari uji asumsi PH untuk masing-masing variabel ditampilkan pada Tabel 4.12.

Tabel 4.12 Hasil uji asumsi PH per variabel

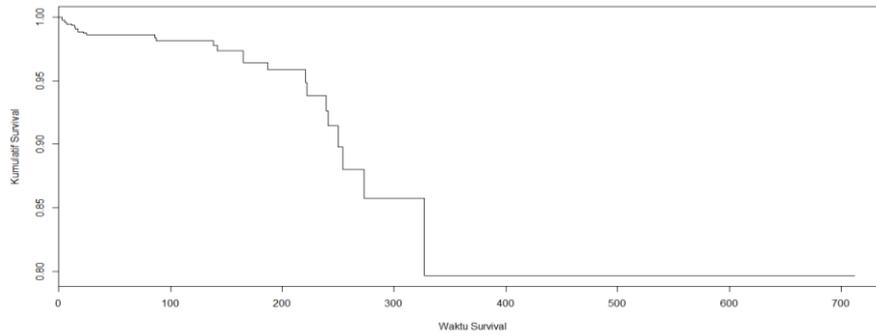
Variabel	Korelasi	p-value
Usia	-0,175	0,391
Komplikasi	-0,269	0,165
Anemia	-0,304	0,087
Jenis pengobatan	0,301	0,096
Stadium	-0,444	0,031
Umur saat menikah pertama	0,241	0,139
Umur saat menstruasi pertama	0,085	0,693
Siklus menstruasi	0,228	0,209
Lama menstruasi	0,067	0,714
Paritas	0,663	0,001
Jenis KB	-0,087	0,612
Tingkat pendidikan	-0,157	0,331

Berdasarkan hasil pada Tabel 4.12 diketahui bahwa variabel stadium dan paritas tolak H_0 artinya ada korelasi antara Schoenfeld residual dan ranking *survival time* sehingga asumsi PH tidak terpenuhi untuk variabel tersebut. Karena stadium

adalah variabel yang paling penting dalam kanker maka diperlukan model yang tidak memerlukan asumsi PH sebagai pembanding.

4.4.2 Fungsi *Survival* Kumulatif dan Fungsi Hazard Kumulatif

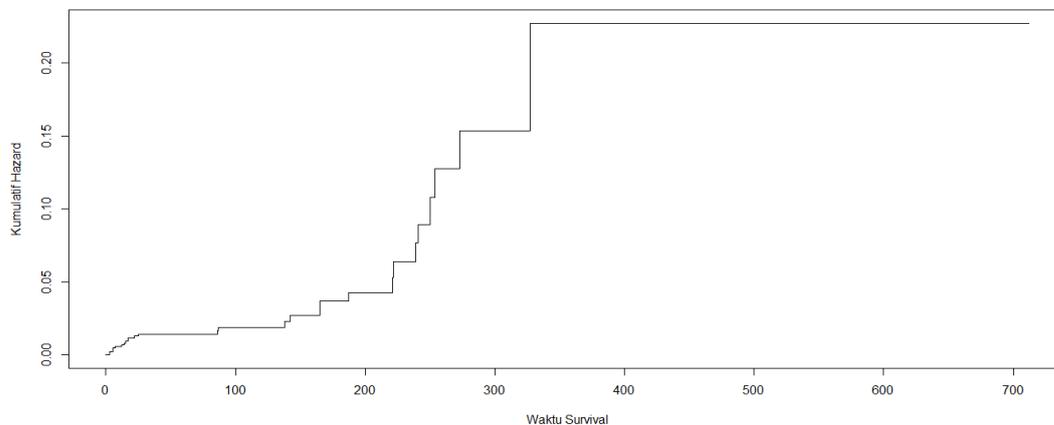
Fungsi *survival* digunakan untuk mengetahui bagaimana probabilitas hidup pasien kanker serviks. Gambar 4.19 adalah grafik dari fungsi *survival* kumulatif.



Gambar 4.19 Kurva *survival* kumulatif

Berdasarkan Gambar 4.19 tampak bahwa semakin lama probabilitas bertahan pasien kanker serviks semakin berkurang pada hari ke-0 sampai dengan hari ke-310, setelah hari ke-310 cenderung konstan.

Fungsi hazard digunakan untuk melihat bagaimana laju pasien untuk mengalami *event* (dalam hal ini meninggal) saat *survival time* semakin panjang. Gambar 4.20 menjelaskan grafik dari fungsi hazard kumulatif.



Gambar 4.20 Kurva *hazard* kumulatif

Berdasarkan Gambar 4.20 tampak bahwa semakin panjang *survival time* semakin cepat laju pasien untuk mengalami *event*, hal ini dapat dilihat dari semakin naiknya kurva fungsi hazard. Jika dihubungkan dengan fungsi *survival*, semakin cepat laju

seorang pasien untuk mengalami *event*, maka semakin kecil pula peluangnya untuk bertahan hidup.

4.4.3 Penerapan Model Cox PHM pada Data Pasien Kanker Serviks

Model Cox PHM merupakan model yang paling sering digunakan untuk menganalisis data *survival* yang memiliki *survival time* dan status tersensor. Berdasarkan data pasien kanker serviks yang digunakan dalam penelitian ini, diperoleh hasil estimasi parameter masing-masing variabel yang disajikan pada Tabel 4.13.

Tabel 4.13 Hasil estimasi parameter model Cox PHM

Variabel	$\hat{\beta}$	SE	Z	p-value	HR
Usia	0,008	0,026	0,31	0,760	1,008
Komplikasi	0,188	0,608	0,31	0,758	1,207
Anemia	0,939	0,713	1,32	0,188	2,557
Pengobatan 0 (Kemoterapi)	-0,559	0,734	-0,76	0,446	0,571
Pengobatan 1 (Transfusi PRC)	0,256	0,772	0,33	0,741	1,291
Pengobatan 2 (Kemoterapi dan Transfusi PRC)	-2,095	0,743	-2,82	0,005	0,123
Stadium	2,797	0,719	3,89	0,000	16,389
Umur menikah pertama	0,030	0,054	0,55	0,579	1,031
Umur menstruasi pertama	-0,083	0,144	-0,58	0,564	0,920
Siklus menstruasi	0,287	0,219	1,3	0,192	1,332
Lama menstruasi	-0,191	0,177	-1,08	0,281	0,826
Paritas	-0,071	0,153	-0,47	0,641	0,931
KB 0 (Tidak KB)	0,463	0,686	0,67	0,499	1,589
KB 1 (KB suntik)	-0,137	0,774	-0,18	0,859	0,872
KB 2 (KB pil)	-0,930	1,009	-0,92	0,356	0,394
Pendidikan 0 (SD)	-1,336	0,576	-2,32	0,020	0,263
Pendidikan 1 (SMP)	-0,321	0,685	-0,47	0,639	0,725
Likelihood Ratio Test	47,8	df=17		p-value=0,000	

Pengujian parameter secara serentak menggunakan Likelihood Ratio test dimana pada Tabel 4.13 diketahui bahwa nilai p-value dari Likelihood Ratio-nya adalah 0,000, artinya minimal ada satu variabel yang berpengaruh secara signifikan pada laju kematian pasien. Selanjutnya untuk hasil pengujian parsial pada Tabel 4.13 diketahui bahwa variabel yang signifikan mempengaruhi laju kematian pasien

kanker serviks adalah pengobatan 2 (Kemoterapi dan Transfusi PRC), stadium, dan pendidikan 0. Model Cox PHM yang terbentuk adalah sebagai berikut:

$$\hat{h}(t) = h_0(t) \exp[0,008(\text{usia}) + 0,188(\text{komplikasi}) + 0,939(\text{anemia}) - 0,559(\text{pengobatan0}) + 0,256(\text{pengobatan 1}) - 2,095(\text{pengobatan 2}) - 2,797(\text{stadium}) + 0,030(\text{umur menikah pertama}) - 0,083(\text{umur menstruasi pertama}) + 0,287(\text{siklus menstruasi}) - 0,191(\text{lama menstruasi}) - 0,071(\text{paritas}) + 0,463(\text{KB0}) - 0,137(\text{KB1}) - 0,930(\text{KB2}) - 1,336(\text{pendidikan 0}) - 0,321(\text{pendidikan 2})].$$

Dalam model Cox PHM, HR (hazard ratio) digunakan untuk menginterpretasikan variabel prediktor terhadap laju kematian pasien kanker serviks. Misalkan untuk variabel usia diperoleh nilai HR sebesar 1,008 artinya setiap penambahan usia pasien 1 tahun akan meningkatkan laju kematian pasien kanker serviks sebesar 1,008 kali. Nilai HR pada variabel pengobatan 2 adalah 0,123 artinya pasien yang menjalani pengobatan 2 memiliki laju kematian 0,123 kali daripada pasien yang menjalani pengobatan 3. Nilai HR variabel stadium 16,389 artinya pasien dengan kode stadium1 (stadium3 dan stadium 4) memiliki laju kematian 16,389 kali daripada pasien dengan kode stadium0 (stadium1 dan stadium 2). Nilai HR pendidikan0 adalah 0,263, artinya pasien dengan pendidikan0 memiliki laju kematian 0,263 kali daripada pasien dengan pendidikan2. Interpretasi dengan cara yang sama juga berlaku untuk variabel yang lain.

Performansi model Cox PHM berdasarkan *prognostic index* dapat dilihat berdasarkan nilai hazard ratio, log-rank, dan c-index pada data pasien kanker serviks yang ditampilkan pada Tabel 4.14.

Tabel 4.14 Performansi model Cox PHM pada data pasien kanker serviks

Hazard Ratio (HR)	Log-Rank (LR)	c-index
0,147	16,661	11,217%

Tabel 4.14 menampilkan performansi model Cox PHM dimana pada penentuan HR dan LR didasarkan pada pengelompokan pasien berdasarkan nilai prognosticnya. Dalam hal ini *prognostic index* menggantikan bersesuaian dengan fungsi hazard pada model Cox PHM, yaitu saat *prognostic index* tinggi maka fungsi hazard juga tinggi. Berdasarkan nilai *prognostic index*, pasien dikelompokkan

menjadi *high risk* dan *low risk*. HR dan LR digunakan untuk melihat apakah ada perbedaan kurva *survival time* antara pasien yang masuk dalam kelompok *high risk* dan pasien yang masuk dalam kelompok *low risk*. Berdasarkan nilai HR, terlihat perbedaan antara pasien yang masuk dalam *high risk* dan *low risk* karena nilai dari HR yang jauh dari 1. Hal ini diperkuat dengan hasil pada LR yang menggunakan uji chi-square dengan $df = 1$ dan $\alpha = 0,01$ diperoleh hasil menolak H_0 artinya secara statistik terdapat perbedaan kurva *survival* yang signifikan antar dua kelompok tersebut. Performansi lain adalah c-index yang dapat mengukur keterurutan antara *survival time* observasi dengan prognostic prediksi. Model Cox PHM pada kasus ini mempunyai c-index 11,217%, artinya model ini menghasilkan keterurutan prognostic dan *survival time* yang sesuai sebesar 11,217%.

4.4.4 Penerapan Model SURLS-SVM pada data pasien kanker serviks

Selain menggunakan pendekatan klasik yaitu Cox PHM, data pasien kanker serviks juga akan dianalisa menggunakan pendekatan *machine learning* yaitu SURLS-SVM. Model ini memerlukan optimasi parameter, yaitu parameter SURLS-SVM (γ) dan parameter kernel (σ^2) dalam hal ini menggunakan kernel RBF. Penentuan nilai parameter yang optimal dilakukan dengan metode yang umum digunakan yaitu *gridsearch*. Parameter yang optimal diukur berdasarkan nilai c-index yang dihasilkan dari setiap iterasi. Hasil performansi untuk masing-masing iterasi ditampilkan pada Lampiran 10. Dari hasil iterasi diperoleh nilai c-index tertinggi untuk SURLS-SVM diperoleh saat $C = 0,5$ dan $\gamma = 1$. Performansi dari SURLS-SVM pada data pasien kanker serviks ditampilkan pada Tabel 4.15.

Tabel 4.15 Performansi model SURLS-SVM pada data pasien kanker serviks

Hazard Ratio (HR)	Log-Rank (LR)	c-index
$9,12 \times 10^{-11}$	324,799	96,447%

Pada Tabel 4.15 dapat dilihat bahwa nilai c-index dari model SURLS-SVM adalah 96,447%, artinya model ini dapat memprediksi prognostic lebih baik daripada Cox PHM sehingga nilai keterurutan (*concordance*) antara *survival time* dan prediksi prognostic lebih tinggi dibandingkan model klasik. Sama seperti pada model Cox PHM, model SURLS-SVM juga menganalisis perbedaan kurva *survival time* antara pasien yang masuk dalam kelompok *high risk* dan pasien yang masuk

dalam kelompok *low risk* berdasarkan nilai *prognostic index*. Perbedaan kurva *survival time* untuk kedua kelompok dapat dilihat berdasarkan nilai HR dan LR. Dari Tabel 4.15 diketahui bahwa nilai HR jauh dari 1 artinya terdapat perbedaan kurva *survival time* antara kedua kelompok pasien, sedangkan dari pendekatan uji statistik dapat dilihat berdasarkan nilai LR yang menggunakan uji chi-square dan diperoleh hasil untuk menolak H_0 artinya terdapat perbedaan kurva *survival time* atau model ini dapat memisahkan pasien yang masuk dalam kelompok *high risk* dan pasien yang masuk dalam kelompok *low risk*.

4.4.5 Penerapan *Feature Selection* pada Data Pasien Kanker Serviks

Pada model SURLS-SVM diperoleh nilai c-index yang lebih baik dari model Cox PHM tetapi belum diketahui variabel yang mempengaruhi nilai c-index. Untuk mengetahui variabel yang berpengaruh terhadap nilai c-index, maka dilakukan *feature selection* dengan metode *backward elimination*. Metode ini dilakukan dengan cara menghapus satu per satu variabel kemudian dihitung selisih antara c-index sebelum variabel tersebut dihapus dengan c-index setelah variabel tersebut dihapus. Langkah ini dilakukan sampai didapatkan nilai c-index terbaik, yaitu ketika sudah tidak ada lagi variabel yang bisa dihapus karena akan mengurangi nilai c-index.

Dengan menggunakan parameter kernel dan parameter SURLS-SVM berturut turut 1 dan 4, diperoleh hasil variabel yang terpilih setelah dilakukan *feature selection* adalah usia (P_1), status komplikasi (P_2), jenis pengobatan (P_4), stadium (P_5), umur pertama kali menikah (P_6), siklus menstruasi (P_8), lama menstruasi (P_9), paritas (P_{10}), dan tingkat pendidikan (P_{12}). Urutan variabel yang keluar terlebih dahulu dari model sampai yang keluar terakhir yaitu status anemia (P_3) yang menaikkan c-index menjadi 97.089%, selanjutnya jenis KB (P_{11}) yang menaikkan c-index menjadi 97.138%, dan variabel yang terakhir keluar adalah umur pertama kali menstruasi (P_7) yang menaikkan c-index menjadi 97.171%. Performansi pada Cox PHM dan SURLS-SVM sesudah dilakukan *feature selection* dapat dilihat pada Tabel 4.16.

Tabel 4.16 Performansi Cox PHM dan SURLS-SVM setelah *feature selection*

Cox PHM			SURLS-SVM		
HR	LR	c-index	HR	LR	c-index
0,123	20,231	11,086%	$7,44 \times 10^{-11}$	277,039	97,171%

Jika dibandingkan sebelum dan setelah dilakukan *feature selection*, c-index dari Cox PHM turun sebesar 1,173% dan untuk SURLS-SVM nilai c-index meningkat sebesar 0,75%. Dari variabel yang terpilih dapat dilihat urutan variabel yang memberikan pengaruh besar pada c-index. Variabel yang jika dihapus dari model menyebabkan nilai c-index menjadi berkurang besar, maka variabel itulah yang memberikan pengaruh besar pada nilai c-index. Berdasarkan hasil yang diperoleh dari model SURLS-SVM, nilai c-index untuk masing-masing variabel ditampilkan pada Tabel 4.17.

Tabel 4.17 Pengaruh variabel terpilih terhadap nilai c-index

Variabel yang dihapus	c-index sebelum variabel dihapus	c-index setelah variabel dihapus	Selisih
P_1	97,171%	96,604%	-0,567
P_2	97,171%	97,171%	0,000
P_4	97,171%	97,138%	-0,033
P_5	97,171%	97,105%	-0,066
P_6	97,171%	96,653%	-0,518
P_8	97,171%	97,122%	-0,049
P_9	97,171%	97,171%	0,000
P_{10}	97,171%	97,155%	-0,016
P_{12}	97,171%	97,171%	0,000

Berdasarkan Tabel 4.17 dapat diketahui bahwa variabel yang menyebabkan c-index turun paling besar jika variabel tersebut dihapus dari model adalah P_1 (usia), artinya variabel usia merupakan faktor terbesar yang mempengaruhi keterurutan antara *survival time* dengan *prognostic index*. Urutan variabel yang paling mempengaruhi c-index setelah variabel P_1 adalah P_5 (stadium), P_{10} (paritas), P_9 (lama menstruasi), P_8 (siklus menstruasi), P_2 (status komplikasi), P_4 (jenis pengobatan), P_6 (umur pertama kali menikah), dan P_{12} (tingkat pendidikan). Jika dihubungkan dengan uji signifikansi parameter pada model Cox PHM, variabel yang sesuai adalah P_4 (jenis pengobatan), P_5 (stadium), dan P_{12} (tingkat

pendidikan), karena pada model Cox PHM tiga variabel ini secara signifikan mempengaruhi laju meninggalnya pasien kanker serviks.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut.

1. *Prognostic index* merupakan pengganti dari fungsi hazard pada Cox PHM yang dapat digunakan untuk menghitung performansi dari SURLS-SVM maupun Cox PHM. Persamaan untuk menghitung *prognostic index* adalah sebagai berikut:

$$\begin{aligned}\hat{\mathbf{u}}^* &= \sum_{i=1}^n \sum_{i < j}^n [\alpha_{ij} (\boldsymbol{\varphi}(\mathbf{x}_i) - \boldsymbol{\varphi}(\mathbf{x}_j))]^T \boldsymbol{\varphi}(\mathbf{x}^*) \\ &= \sum_{i=1}^n \sum_{i < j}^n \alpha_{ij}^T [\boldsymbol{\varphi}(\mathbf{x}_i)^T \boldsymbol{\varphi}(\mathbf{x}^*) - \boldsymbol{\varphi}(\mathbf{x}_j)^T \boldsymbol{\varphi}(\mathbf{x}^*)] \\ &= \boldsymbol{\alpha}^T \mathbf{DK}_n(\mathbf{x}^*)\end{aligned}$$

dengan:

$\boldsymbol{\alpha}$: pengali Lagrange yang merupakan hasil dari optimasi,

\mathbf{D} : matriks yang memuat $\{-1,0,1\}$

$\mathbf{K}_n(\mathbf{x}^*) = [\mathbf{K}(\mathbf{x}_1, \mathbf{x}^*) \dots \mathbf{K}(\mathbf{x}_n, \mathbf{x}^*)]^T$ merupakan matriks Kernel.

2. Aplikasi Cox PHM dan SURLS-SVM pada data simulasi dan data real (data kanker serviks) memberikan hasil bahwa berdasarkan ukuran performansi (c-index, log-rank, dan hazard ratio) SURLS-SVM selalu memberikan hasil yang lebih tinggi dari Cox PHM berdasarkan nilai c-index, untuk dua ukuran performansi yang lain memberikan hasil yang kurang konsisten karena saat nilai hazard ratio tinggi maka nilai log-rank rendah begitu juga sebaliknya, sehingga c-index merupakan ukuran yang paling konsisten. Persentase tersensor memberikan efek yang paling besar terhadap perubahan performansi SURLS-SVM, semakin besar persentase tersensor maka semakin besar pula nilai c-index.
3. Penerapan *feature selection (backward selection)* pada model SURLS-SVM mampu menaikkan nilai c-index dan mengeliminasi *feature* yang tidak

signifikan. Dari hasil *feature selection* diperoleh variabel yang mempengaruhi menurut model SURLS-SVM adalah usia, status komplikasi, jenis pengobatan (kemoterapi, transfusi PRC, lainnya), stadium, usia saat menikah pertama, siklus menstruasi, lama menstruasi, paritas, dan tingkat pendidikan. Sedangkan berdasarkan hasil Cox PHM, *feature* yang berpengaruh secara signifikan adalah jenis pengobatan, stadium, dan tingkat pendidikan dimana tiga *feature* ini juga masuk dalam model SURLS-SVM.

5.2 Saran

Beberapa saran yang dapat penulis berikan dengan harapan dapat dikembangkan untuk penelitian selanjutnya adalah sebagai berikut.

1. Jika asumsi proporsional hazard dilanggar, digunakan sebagai alternatif yaitu Straified Cox PHM. Hasil dari Stratified Cox PHM dapat dibandingkan dengan SURLS-SVM.
2. Pada penelitian ini interaksi antar variabel yang digunakan untuk membangkitkan *survival time* pada data simulasi non-linier tidak dimasukkan sebagai prediktor saat analisis model, sehingga hasil yang diperoleh saat menghitung persentase signifikansi masing-masing model menjadi kurang sesuai. Pada penelitian selanjutnya, interaksi ini selain digunakan untuk membangkitkan *survival time* juga dapat digunakan sebagai *feature* yang masuk dalam model sehingga hasilnya lebih mudah untuk diinterpretasikan.
3. *Feature selection* yang digunakan dalam penelitian ini adalah metode *backward elimination* yang masih sangat sederhana sehingga kurang sensitif terhadap variabel yang tidak signifikan. Pada penelitian selanjutnya dapat digunakan metode *feature selection* yang lain sehingga variabel yang tidak signifikan tidak masuk dalam model akhir.

Daftar Pustaka

- Abadi, A. (2011), "Comparison of Aalen's Additive and Cox Proportional Hazard Models for Breast Cancer Survival: Analysis of Population-Based Data from British Columbia, Canada", *Asian Pacific Journal of Cancer Prevention*, 12,3113-3116.
- Afifah, A.N. dan Purnami, S.W. (2016), "Uji Proportional Hazard pada Data Penderita Kanker Serviks di RSUD dr. Soetomo Surabaya", *Jurnal Sains dan Seni ITS Vol.5 No.1*.
- Bender, R., Augustin, T. dan Blettner, M. (2005), "Generating survival times to simulate Cox proportional hazards models", *Statistics in Medicine*, 24(11) : 1713-1723.
- Chandrashekar, G. dan Sahin,F. (2013), "A survey on feature selection methods", *Computers and Electrical Engineering*, 40: 16-28.
- Collett, D. (2003), *Modelling Survival Data in Medical Research* (2nd ed.), Chapman and Hall, London.
- Diananda, R. (2007), *Mengenal Seluk Beluk Kanker*, Katahati, Yogyakarta.
- Dzion, D.S., Krycman, M.L. dan Disilvetto, P.A. (2011), *100 Tanya Jawab Mengenai kanker Serviks*, PT. Indeks, Jakarta.
- Endo, D., Todo, Y., Okamoto, K., Minobe, S., Kato, H., dan Nishiyama, N., (2015), "Prognostic Factors for patients with cervical cancer treated with concurrent chemoradiotherapy: a retrospective analysis in Japan cohort", *Journal of Gynecologic Oncology*, Vol 26 pp.12-16
- Frank, E., Harrell, JR. dan Kerry, L.L. (1984), "Regression Modelling Strategies for Improved Prognostic Prediction", *Statistics in Medicine*, Vol.3 pp.143-152.
- Furqon, A. (2013), *Analisis Regresi Weibull untuk Mengetahui Faktor-Faktor yang Mempengaruhi Laju Perbaikan Klinis Penderita Stroke*, Tugas Akhir Jurusan Statistika ITS.
- Galea, M.H., Blamey, R.W., Elston, C.E. dan Ellis, I.O. (1992), "The Nottingham Prognostic Index in Primary Breast Cancer", *Breast Cancer Research and Treatment* 22 : 207-219, Kluwer Academic Publishers, Netherlands.
- Gunn, S.R. (1998), *Support Vector Machine for Classification and Regression, Technical Support*, University of Southampton, Southampton.

- Goli, S. Mahjub, and Faradmal, J. (2016), "Survival Prediction and Feature Selection in Patients with Breast Cancer Using Support Vector Regression", *Computational and Mathematical Methods in Medicine*.
- Haerdle, W.K., Prastyo, D.D. dan Hafner, C.M. (2014), "Support Vector Machines with Evolutionary Model Selection for Default Prediction", dalam Racine, J., Su, L. dan Ullah, A. (Eds), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, New York, 346-373.
- Howley, P. dan Lowy, D. (2001), *Papillomaviruses and Their Replication*, *Fields Virology* ed.4th pp (2197-2209), Lippincott-Raven, Philadelphia (PA).
- Inayati, K.D. dan 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 Vol.4 No.2*.
- Ioka, A., Tsukuma, H., Ajiki, W. dan Oshima, A. (2005), "Influence of Age on Cervical Cancer Survival in Japan", *Jpn J Clin Oncol*, 35 (8), 464-469.
- Kementrian Kesehatan RI. (2015), *Panduan Pelayanan Klinik Kanker Serviks Komisi Penanggulangan Kanker Nasional*.
- Khaulasari, H. (2016), *Combine Sampling-Least Square Support Vector Machine untuk Klasifikasi Multi Class Imbalanced Data*, Tesis Jurusan Statistika.
- Kleinbaum, D.G. dan Klein, M. (2012), *Survival Analysis : A Self-Learning Text* (Third ed.), Springer, London.
- Kriner, M. (2007), *Survival Analysis with Multivariate Adaptive Regression Splines*, Disertasi Universität München.
- Kustiyati, S. dan Winarni, (2011), *Deteksi Dini Kanker Leher Rahim dengan Metode IVA di Wilayah Kerja Puskesmas Ngoresan Surakarta*, *Gaster* Vol.8 No.1 (681-694).
- Lee, E.T. dan Wang, J.W. (2003), *Statistical Methods for Survival Data Analysis* (Third ed.), John Wiley and Sons.Inc, New York.
- Lee, Y.Y. dan Kim, T.J. (2013), "Genetic Profiling to Predict Recurrence of Early Cervical Cancer", (Elsevier, Penyunt.) *Gynecology Oncology*, 131, 650-654.
- Lestari, H.N. (2016), *Analisis Survival Laju Perbaikan Klinis Pasien Penyakit Jantung Koroner di RSUD dr.Soetomo Surabaya dengan Pendekatan Multi Period Logit*, Tugas Akhir Jurusan Statistika ITS.

- Mahjub, H., Faradmal, J., Soltanian, A.R. dan Goli, S. (2016), "Performance Evaluation of Support Vector Regression Models for Survival Analysis : A Simulation Study", *International Journal of Advanced Computer Science and Applications*, Vol.7 No.6.
- Melva. (2008), *Faktor-Faktor yang Mempengaruhi Kejadian Kanker Leher Rahim pada Penderita yang Datang Berobat di RSUP H.Adam Malik Medan Tahun 2008*, Tesis Program Magister Administrasi dan Kebijakan Kesehatan Universitas Sumatera Utara.
- Nisa, A.F. dan Budiantara, I.N. (2013), "Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD)", *Jurnal Sains dan Seni ITS*.
- Ocviyanti, D. dan Handoko, Y. (2013), "Peran Dokter Umum dalam Pencegahan Kanker Serviks di Indonesia", *J Indon Med Assoc*, 63(1),1-3.
- Prawirohardjo. (2001) , *Ilmu Kandungan*, YBPSP, Jakarta.
- Purnami, S.W., Inayati, K.D., Sari, N.W.W., Chosuvivatwong, V. dan Sriplung, H. (2015), "Survival Analysis of Cervical Cancer using Stratified Cox Regression", *Symposium on Biomathematics*.
- Rahma, G. (2012), "*Hubungan Jaminan Pembiayaan Asuransi Kesehatan dengan Ketahanan Hidup Lima Tahun Pasien Kanker Serviks di Rumah Sakit Umum Pusat Nasional Dr. Cipto Mangunkusumo Jakarta Tahun 2007-2010*", Skripsi Fakultas Kesehatan Masyarakat UI.
- 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 Jurusan Statistika ITS.
- Samadi, P.H. (2010), *Yes, I know everything about Kanker Serviks*, Tiga Kelana, Yogyakarta.
- Sirait, A.M., Soetiarto, F. dan Oemiati, R. (2003), "Ketahanan Hidup Penderita Kanker Serviks di Rumah Sakit Kanker Dharmais Jakarta", *Buletin Penelitian Kesehatan*, Vol.31, No.1 : 13-24.
- Suykens, J.A.K. dan Vandewalle, J. (1999), "Least Squares Support Vector Machine Classifiers", *Neural Processing Letters*, 9, 293-300.
- Van Belle, V., Pelckmans, K., Suykens, J.A.K., dan Van Huffel, S. (2010a), "Additive Least Squares Support Vector Machines", *Statistics in Medicine*, 29(2) : 296-308.

- Van Belle, V., Pelckmans, K., Suykens, J.A.K., dan Van Huffel, S. (2010b), "Improved performance on high-dimensional survival data by application of Survival-SVM", *Bioinformatics* Vol.27 no.1 (87-94), Oxford University Press.
- Van Belle, V., Pelckmans, K., Suykens, J.A.K., dan Van Huffel, S. (2011), "Support vector methods for survival analysis: a comparison between ranking and regression approaches", *Artificial Intelligence in Medicine*, 53 pp 107-118.
- Vapnik, V. (1998), *Statistical Learning Theory*, Wiley and Sons.
- Wijaya, A. (2015), *Analisis Survival pada Pasien Penderita Sindrom Koroner Akut Di RSUD Dr.Soetomo Surabaya Tahun 2013 Menggunakan Regresi Cox Proportional Hazard*, Tugas Akhir Jurusan Statistika FMIPA ITS, Surabaya.
- WHO, I.A. (2005), *Cervix Cancer Screening* (Vol.10), IARCH Press, Lyon.

LAMPIRAN

LAMPIRAN 1: Sertifikat Laik Etik Pengambilan Data

F.LITB.003



KOMITE ETIK PENELITIAN KESEHATAN
RSUD Dr. SOETOMO SURABAYA
KETERANGAN KELAIKAN ETIK
("ETHICAL CLEARANCE")
377 / 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 :

**" Analisis Data Survival dengan Survival Least Square-SVM
(Studi Simulasi dan Studi Kasus Pasien Kanker Serviks) "**

PENELITI UTAMA : Halwa Annisa Khoiri

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

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

DINYATAKAN LAIK ETIK

SURABAYA, 30 MAY 2017
KETUA



(Dr. Erzeus Hamudito, dr., Sp.An, KIC.KAP)
NIP. 19511007 197903 1 002

LAMPIRAN 2. Data *Survival* Pasien Kanker Serviks di RSUD dr. Soetomo Surabaya

No	T	delta	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	3	0	44	0	0	0	3	18	14	28	7	2	1	0
2	3	0	53	1	0	0	2	0	15	28	7	0	0	2
3	3	1	46	1	1	0	3	16	14	30	5	0	0	2
4	3	0	39	0	0	0	3	20	13	28	4	2	0	1
5	3	0	42	0	1	0	3	20	12	28	7	4	3	0
6	3	1	43	1	1	0	3	18	13	28	7	2	1	2
7	3	0	57	0	0	0	3	18	13	28	7	1	0	1
8	4	0	60	1	1	1	3	18	13	28	7	3	3	2
9	5	1	44	1	0	0	3	19	12	28	7	2	3	2
.
.
.
.
.
404	347	0	56	1	1	2	2	25	13	28	5	3	3	2
405	361	0	48	1	1	2	3	18	14	28	5	2	0	0
406	428	0	54	1	1	0	2	20	9	28	7	1	0	0
407	429	0	46	1	1	2	3	19	10	28	7	3	1	2
408	436	0	51	1	1	2	2	18	17	30	7	5	2	0
409	445	0	65	1	1	2	2	13	12	28	5	3	0	0
410	505	0	32	1	1	2	2	23	12	28	7	3	0	0
411	625	0	49	1	1	2	2	18	13	30	7	2	1	2
412	712	0	43	0	1	3	1	20	14	28	7	3	0	2

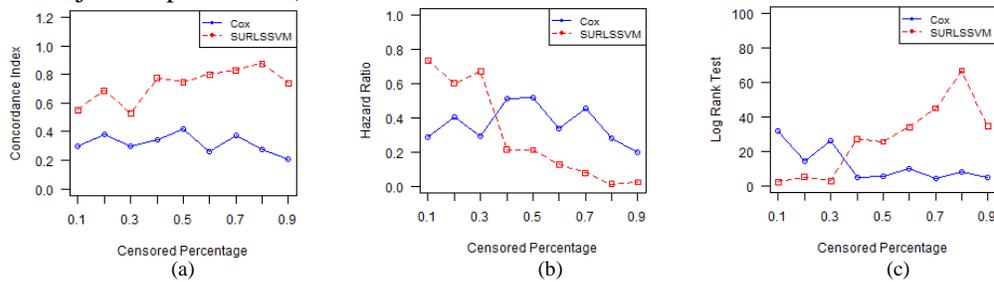
Lampiran 3.a. Tabulasi silang antara status tersensor dan stadium

	Stadium 1	Stadium 2	Stadium 3	Stadium 4	Total
Bertahan	17 4.13%	171 41.5%	192 46.6%	5 1.21%	385 93.45%
<i>Expected Value</i>	15.9	164.5	200	4.7	385
Meninggal	0 0	4 0.97%	22 5.34%	1 0.24%	27 6.55%
<i>Expected Value</i>	1.1	11.5	14	0.3	27
Total	17 4.13%	175 42.48%	214 51.94%	6 1.46%	412 100%

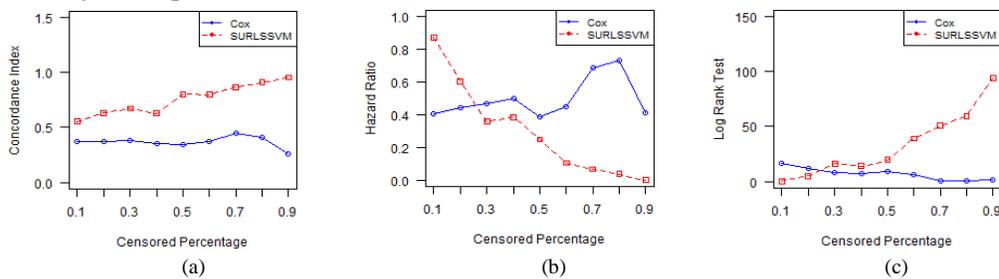
Lampiran 3.b. Tabulasi silang antara status tersensor dan tingkat pendidikan

	SD	SMP	SMA	Perguruan Tinggi	Total
Bertahan	121 29.37%	60 14.56%	198 48.06%	6 1.46%	385 93.45%
<i>Expected Value</i>	117.7	58.9	202.8	5.6	385
Meninggal	5 1.21%	3 0.73%	19 4.61%	0 0	27 6.55%
<i>Expected Value</i>	8.3	4.1	14.2	0.4	27
Total	126 30.58%	63 15.29%	217 52.67%	6 1.46%	412 100%

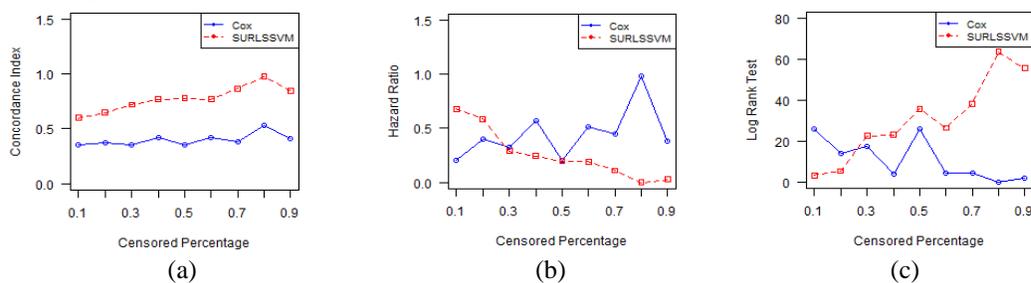
LAMPIRAN 4.a. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 linier, jumlah prediktor 6)



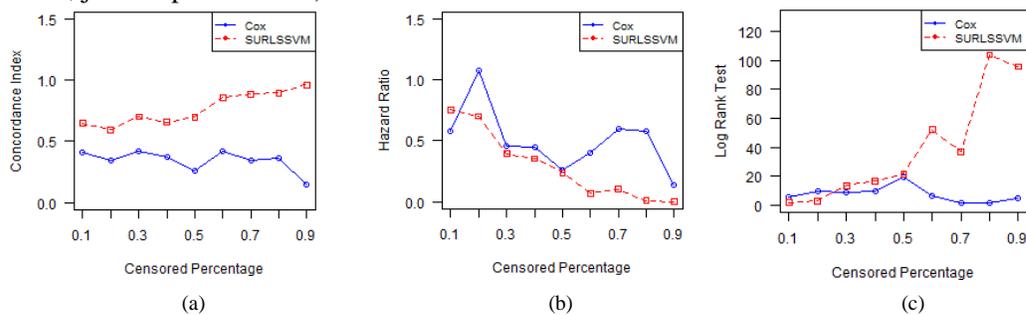
LAMPIRAN 4.b. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 linier, jumlah prediktor 8)



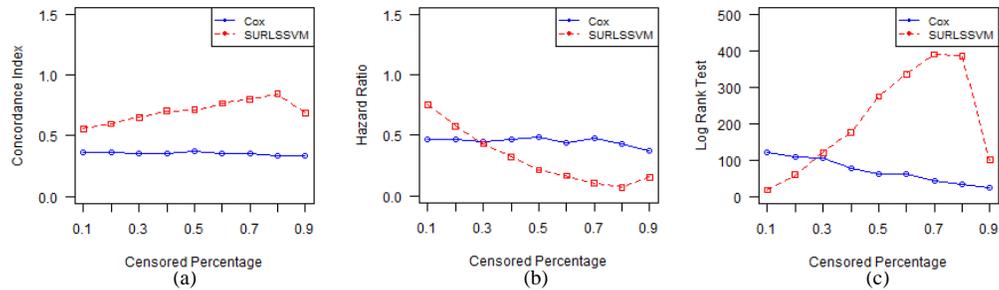
LAMPIRAN 4.c. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 linier, jumlah prediktor 15)



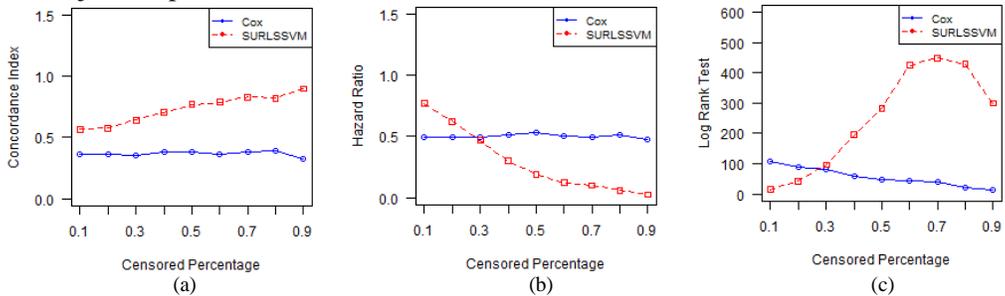
LAMPIRAN 4.d. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 linier, jumlah prediktor 17)



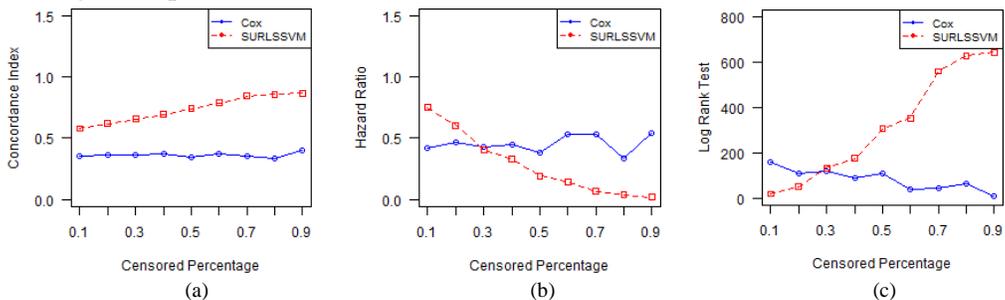
LAMPIRAN 5.a. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 linier, jumlah prediktor 6)



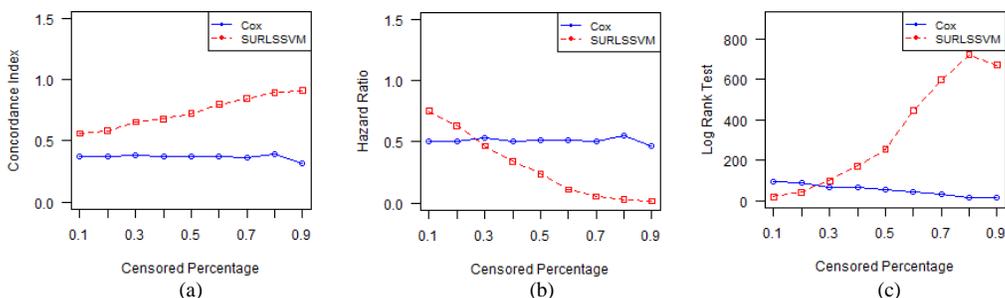
LAMPIRAN 5.b. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 linier, jumlah prediktor 8)



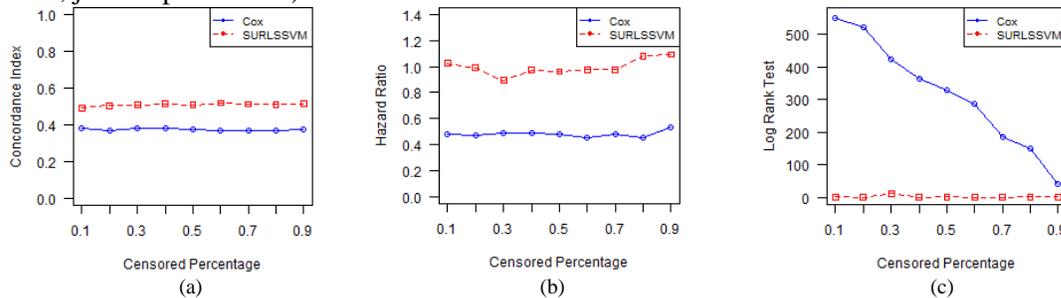
LAMPIRAN 5.c. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 linier, jumlah prediktor 15)



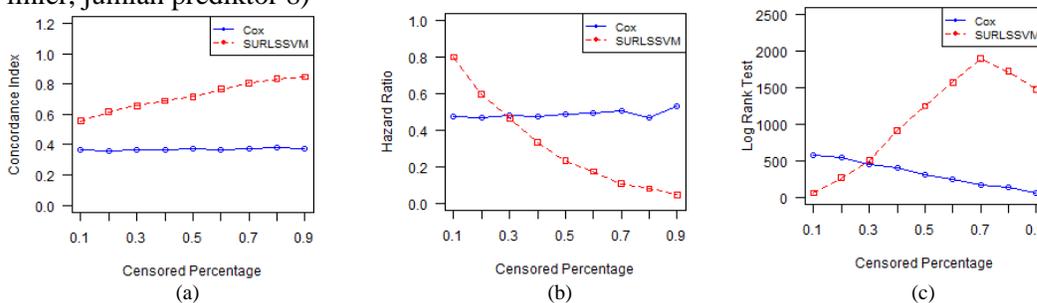
LAMPIRAN 5.d. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 linier, jumlah prediktor 17)



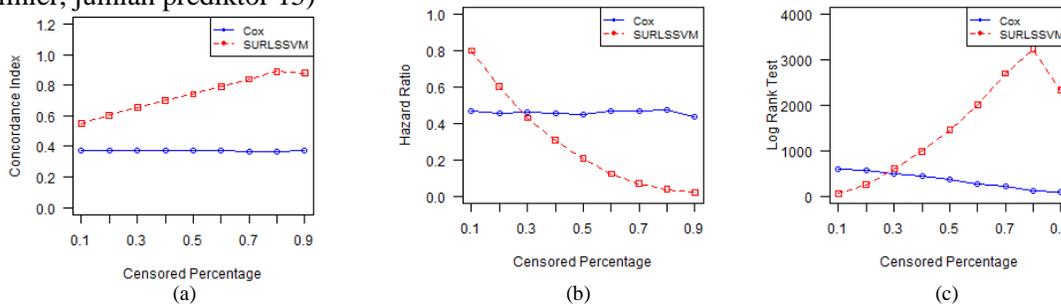
LAMPIRAN 6.a. Ukuran Performansi Berdasarkan Persentase Tersensor (*sample size 5000*, linier, jumlah prediktor 6)



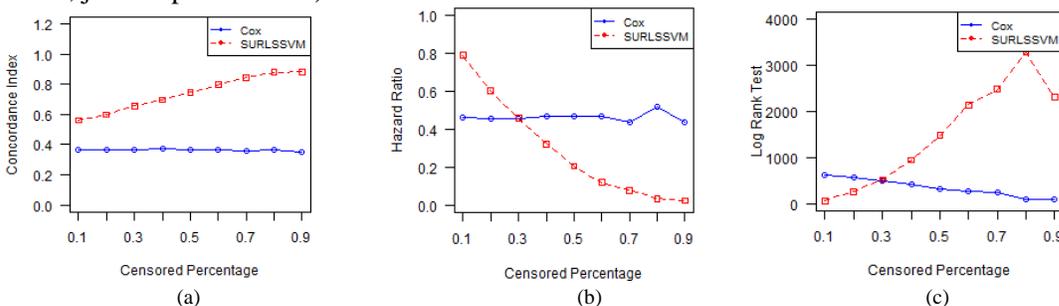
LAMPIRAN 6.b. Ukuran Performansi Berdasarkan Persentase Tersensor (*sample size 5000*, linier, jumlah prediktor 8)



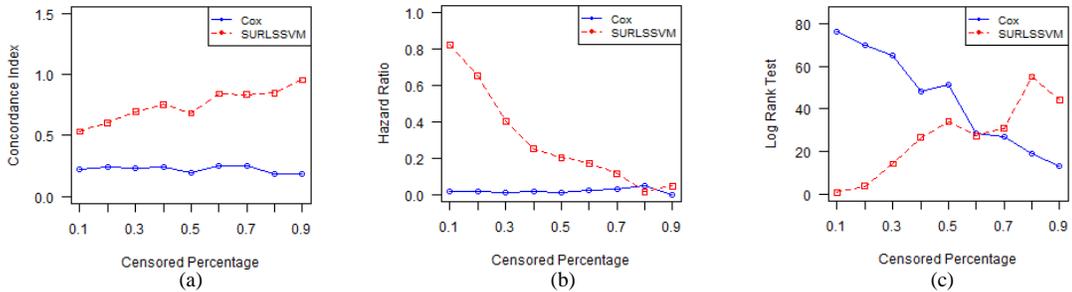
LAMPIRAN 6.c. Ukuran Performansi Berdasarkan Persentase Tersensor (*sample size 5000*, linier, jumlah prediktor 15)



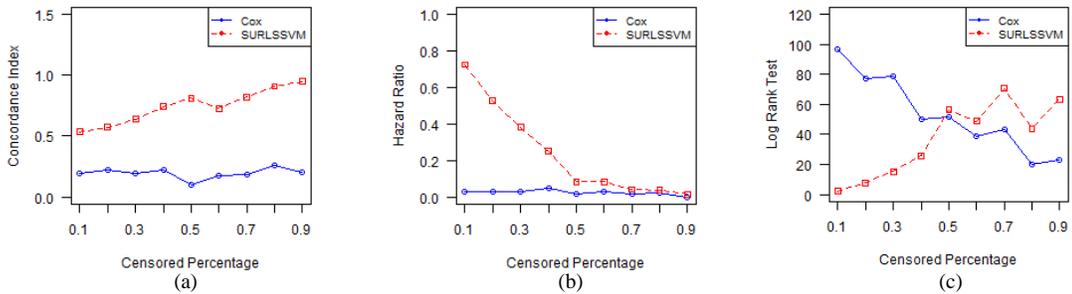
LAMPIRAN 6.d. Ukuran Performansi Berdasarkan Persentase Tersensor (*sample size 5000*, linier, jumlah prediktor 17)



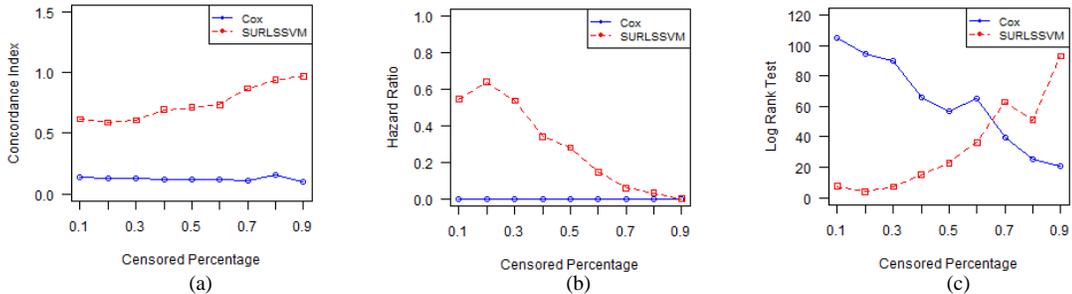
LAMPIRAN 7.a. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 non-linier, jumlah prediktor 6)



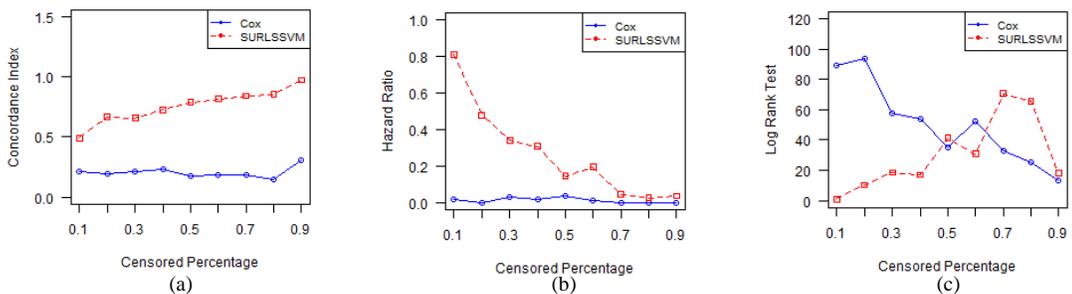
LAMPIRAN 7.b. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 non-linier, jumlah prediktor 8)



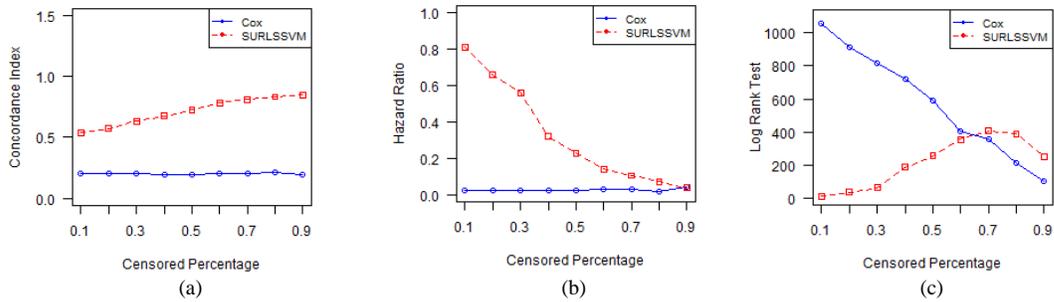
LAMPIRAN 7.c. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 non-linier, jumlah prediktor 15)



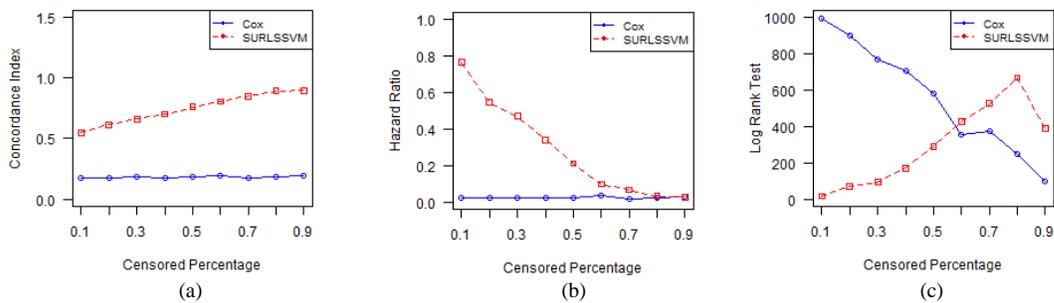
LAMPIRAN 7.d. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 100 non-linier, jumlah prediktor 17)



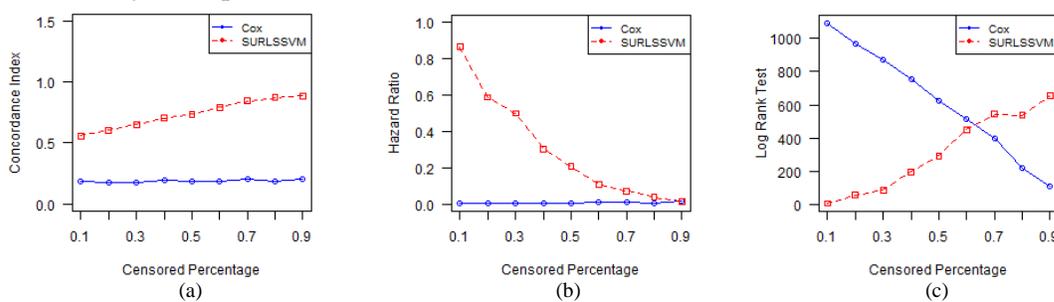
LAMPIRAN 8.a. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 non-linier, jumlah prediktor 6)



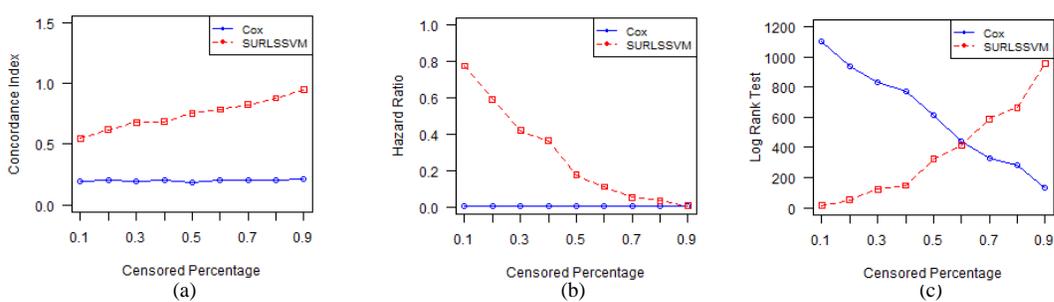
LAMPIRAN 8.b. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 non-linier, jumlah prediktor 8)



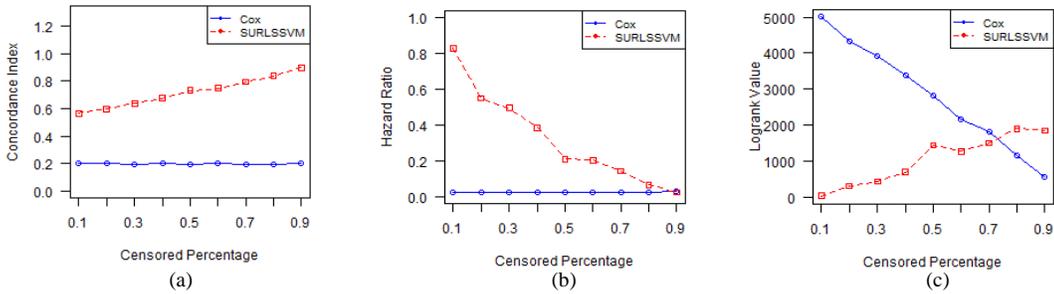
LAMPIRAN 8.c. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 non-linier, jumlah prediktor 15)



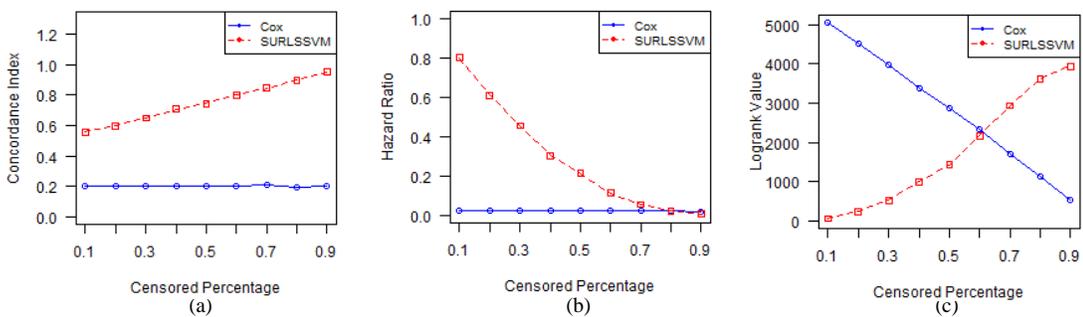
LAMPIRAN 8.d. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 1000 non-linier, jumlah prediktor 17)



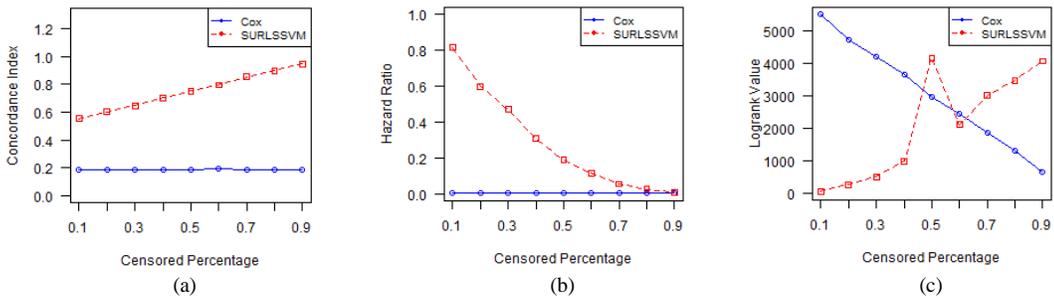
LAMPIRAN 9.a. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 5000 non-linier, jumlah prediktor 6)



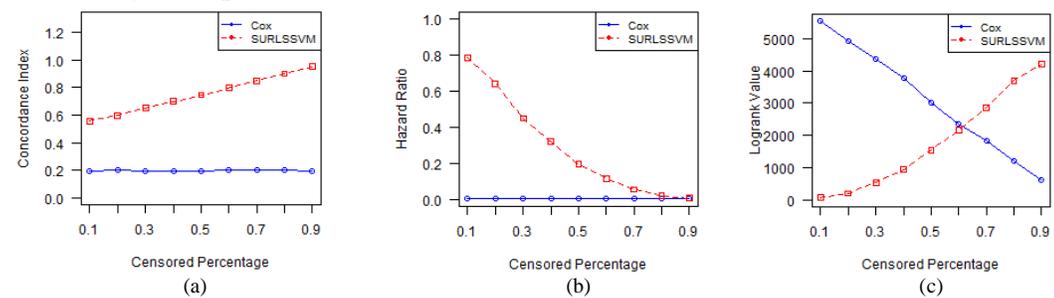
LAMPIRAN 9.b. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 5000 non-linier, jumlah prediktor 8)



LAMPIRAN 9.c. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 5000 non-linier, jumlah prediktor 15)



LAMPIRAN 9.d. Ukuran performansi Berdasarkan Persentase Tersensor (*sample size* 5000 non-linier, jumlah prediktor 17)



LAMPIRAN 10. Iterasi Parameter untuk Model SURLS-SVM pada Data Pasien Kanker Serviks berdasarkan nilai c-index

Iterasi	C	γ	HR	LR	c-index
1	0.0625	0.01	7.68×10^{-11}	329.353	96.299
2	0.0625	0.1	7.68×10^{-11}	329.353	96.299
3	0.0625	1	7.68×10^{-11}	329.353	96.431
4	0.125	0.01	7.68×10^{-11}	329.353	96.299
5	0.125	0.1	7.68×10^{-11}	329.353	96.299
6	0.125	1	7.68×10^{-11}	329.353	96.431
7	0.25	0.01	7.68×10^{-11}	329.353	96.299
8	0.25	0.1	9.41×10^{-11}	323.969	96.266
9	0.25	1	7.68×10^{-11}	329.353	96.431
10	0.5	0.01	0.0059	280.431	96.168
11	0.5	0.1	9.12×10^{-11}	324.799	96.234
12	0.5	1	9.12×10^{-11}	324.799	96.447
13	1	0.01	0.0033	273.867	95.938
14	1	0.1	7.67×10^{-11}	276.207	95.921
15	1	1	9.14×10^{-11}	298.159	96.398
16	2	0.01	0.0124	142.709	92.747
17	2	0.1	1.22×10^{-10}	156.287	93.898
18	2	1	1.20×10^{-10}	183.740	96.382
19	4	0.01	0.0988	33.066	73.487
20	4	0.1	0.0126	72.143	89.046
21	4	1	2.26×10^{-10}	111.383	96.25

LAMPIRAN 11. Syntax R untuk Kurva Kaplan Meier dan Uji Log-Rank

```
library(survival)
library (splines)
data=read.table("E:/data tesis/serviksurut.txt", header = TRUE)
Time = data$T
status = data$status
Y=Surv(Time,status==1)

#Kurva Kaplan Meier tanpa melibatkan faktor
summary(data)
serviks1=data.frame(data)
data
summary(survfit(Surv(data$T,data$status==1)~1))
kmfit1=survfit(Y~1)
summary(kmfit1)
plot(kmfit1, conf.int="none", xlab="Survival Time dalam hari", ylab="Peluang Survival",col=c("red"))

#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X1 (Usia)
umur = data$usia
n =412
rata2_umur = mean(umur)
usia = umur
for (i in 1:n)
{
  if (usia[i]<=rata2_umur)
  {usia[i] = 0}
  else
  {usia[i] = 1}
}
grup_usia = cbind(usia[1:n])
fit = survfit(Y~grup_usia)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("29-49", ">49"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_usia)
Logrank
```

LAMPIRAN 11. Syntax Program R untuk Kurva Kaplan Meier dan Uji Log-Rank (lanjutan)

```
#Kurva Kaplan Meier dan Log Rank Test untuk variabel X2 (Komplikasi)
Komplikasi = data$komplikasi
n =412
fit = survfit(Y~Komplikasi)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("Ada","Tidak
Ada"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~Komplikasi)
Logrank

#Kurva Kaplan Meier dan Log Rank Test untuk variabel X3 (Anemia)
Anemia = data$anemia
n =412
fit = survfit(Y~Anemia)
plot (fit, lty = c("solid","solid"),col = c("red","blue"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("Ya","Tidak"),lty=c("solid","solid"),col=c("red","blue"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~Anemia)
Logrank

#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X4 (Jenis Pengobatan)
pengobatan = data$Jenis.Pengobatan
n = 412
fit = survfit(Y~pengobatan)
plot (fit, lty = c("solid","solid","solid","solid"),col = c("red","blue","green","black"), xlab =
"Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("topright",c("Kemoterapi","Transfusi PRC","Kemoterapi & Transfusi PRC",
"Operasi"),lty=c("solid","solid","solid","solid"),col=c("red","blue","green","black"),cex=
0.8)
Logrank = survdiff(Surv(Time,status)~pengobatan)
Logrank
```

LAMPIRAN 11. Syntax Program R untuk Kurva Kaplan Meier dan Uji Log-Rank (lanjutan)

```
#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X5 (Stadium)
Stadium = data$Stadium
fit = survfit(Y~Stadium)
plot (fit, lty = c("solid","solid","solid", "solid"),col = c("red","blue","green","black"), xlab = "Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("bottomright",c("Stadium I","Stadium II","Stadium III", "Stadium IV"),lty=c("solid","solid", "solid", "solid"),col=c("red","blue","green","black"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~Stadium)
Logrank

#Kurva Kaplan Meier dan Log-Rank test untuk variabel X6 (umur saat menikah)
umur_nikah = data$umur.saat.menikah
n =412
rata2_umurnikah = mean(umur_nikah)
usia_nikah = umur_nikah
for (i in 1:n)
{ if (usia_nikah[i]<=rata2_umurnikah)
  {usia_nikah[i] = 0}
  else {usia_nikah[i] = 1}
}
grup_usianikah = cbind(usia_nikah[1:n])
fit = survfit(Y~grup_usianikah)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("topright",c("0-19",">19"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_usianikah)
Logrank

#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X7 (umur saat menstruasi pertama)
umur_mens = data$umur.pertama.kali.menstruasi
n =412
rata2_umurmens = mean(umur_mens)
usia_mens = umur_mens
for (i in 1:n)
{ if (usia_mens[i]<=rata2_umurmens)
  {usia_mens[i] = 0}
  else {usia_mens[i] = 1}
}
grup_usiamens = cbind(usia_mens[1:n])
fit = survfit(Y~grup_usiamens)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("topright",c("9-13",">13"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_usiamens)
Logrank
```

LAMPIRAN 11. Syntax Program R untuk Kurva Kaplan Meier dan Uji Log-Rank (lanjutan)

```
#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X8 (siklus menstruasi)
siklus_mens = data$Siklus.Menstruasi
n =412
rata2_siklusmens = mean(siklus_mens)
siklus_mens = siklus_mens
for (i in 1:n)
{if (siklus_mens[i]<=rata2_siklusmens)
  {siklus_mens[i] = 0}
  else {siklus_mens[i] = 1}
}
grup_siklusmens = cbind(siklus_mens[1:n])
fit = survfit(Y~grup_siklusmens)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("18-28",">28"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_siklusmens)
Logrank

# Kurva Kaplan Meier dan Log-Rank Test untuk variabel X9 (lama menstruasi)
lama_mens = data$Lama.Menstruasi
n =412
rata2_lamamens = mean(lama_mens)
lama_mens = lama_mens
for (i in 1:n)
{if (lama_mens[i]<=rata2_lamamens)
  {lama_mens[i] = 0}
  else {lama_mens[i] = 1}
}
grup_lamamens = cbind(lama_mens[1:n])
fit = survfit(Y~grup_lamamens)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("3-6",">6"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_lamamens)
Logrank
```

LAMPIRAN 11. Syntax Program R untuk Kurva Kaplan Meier dan Uji Log-Rank (lanjutan)

```
#Kurva Kaplan Meier dan Log-Rank Test untuk variabel X10 (Paritas)
paritas = data$Frekuensi.Melahirkan
n = 412
rata2_Paritas = mean(paritas)
Paritas = paritas
for (i in 1:n)
{
if (Paritas[i]<=rata2_Paritas)
{Paritas[i] = 0}
else {Paritas[i] = 1}
}
grup_paritas = cbind(Paritas[1:n])
fit = survfit(Y~grup_paritas)
plot (fit, lty = c("solid","solid"),col = c("blue","red"), xlab = "Survival time dalam hari",
ylab="Peluang Survival",cex=0.5)
legend ("topright",c("0-2",">2"),lty=c("solid","solid"),col=c("blue","red"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~grup_paritas)
Logrank

# Kurva Kaplan Meier dan Log-Rank Test untuk variabel X11 (Jenis KB)
jeniskb = data$Riwayat.KB
fit = survfit(Y~jeniskb)
plot (fit, lty = c("solid","solid","solid","solid"),col = c("blue","red","green","black"), xlab =
"Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("topright",c("Tidak KB","KB suntik", "KB Pil", "KB
IUD"),lty=c("solid","solid","solid","solid"),col=c("blue","red","green","black"),cex=0.8)
Logrank = survdiff(Surv(Time,status)~jeniskb)
Logrank

# Kurva Kaplan Meier dan Log-Rank Test untuk variabel X12 (Tingkat Pendidikan)
pendidikan = data$Pendidikan
fit = survfit(Y~pendidikan)
plot (fit, lty = c("solid","solid","solid","solid"),col = c("blue","red","green","black"), xlab =
"Survival time dalam hari", ylab="Peluang Survival",cex=0.5)
legend ("topright",c("SD","SMP", "SMA", "Pendidikan
Tinggi"),lty=c("solid","solid","solid","solid"),col=c("blue","red","green","black"),cex=0.
8)
Logrank = survdiff(Surv(Time,status)~pendidikan)
Logrank
```

LAMPIRAN 12. Syntax R untuk Menguji Asumsi Proportional Hazard dengan *Goodness of Fit*

```
library(survival)
library (splines)
data=read.table("E:/data tesis/serviksurut.txt", header = TRUE)
Time = data$T
status = data$status
Y=Surv(Time,status==1)
mod1= coxph(Y~usia+Komplikasi+Anemia+Jenis.Pengobatan+Stadium
           +umur.saar.menikah+umur.pertama.kali.menstruasi+Siklus.Menstruasi
           +Lama.Menstruasi+Frekuensi.Melahirkan+Riwayat.KB+Pendidikan,data=data)
cox.zph(mod1,transform=rank)
```

LAMPIRAN 13. Syntax R untuk Kurva Fungsi *Survival* dan Kurva Fungsi Hazard

```
library(survival)
library (splines)
data=read.table("E:/data tesis/serviksurut.txt", header = TRUE)
Time = data$T
status = data$status
Y=Surv(Time,status==1)

#Kurva Fungsi Survival
s = survfit(mod1, data = data, type ="aalen")$surv
plot(survfit(mod1,data=data,type="aalen"),conf.int="none",xlab="Waktu
Survival",ylab="Kumulatif Survival",ymin=0.98)

#Kurva Fungsi Hazard
h = survfit (mod1, data = data , type ="aalen")$cumhaz
plot(survfit(mod1,data=serviks,type="aalen"),fun="cumhaz",conf.int="none",
xlab="Waktu Survival",ylab="Kumulatif Hazard")
```

LAMPIRAN 14. Syntax R untuk Membangkitkan Data dengan Pendekatan Parametrik

```
n = 100 #sample size, n=1000,5000
# Membangkitkan 6 variabel prediktor
#Distribusi untuk jumlah prediktor yang lain dapat dilihat pada Bab 3)
X1 = rbinom(n, 1, 0.5)
X2 = rbinom(n, 1, 0.3)
X3 = t(rmultinom(n, 1, c(0.2,0.3,0.4, 0.1)))
X3dummy = X3[,-1] X31 = X3dummy[,1] X32 = X3dummy [,2] X33 = X3dummy [,3]
X4 = t(rmultinom(n, 1, c(0.5,0.1,0.2,0.2)))
X4dummy = X4[,-1] X41 = X4dummy [,1] X42 = X4dummy [,2] X43 = X4dummy [,3]
X5 = rnorm(n,40,3)
X6 = rnorm(n, 25, 2)
# Membangkitkan fungsi survival(berupa peluang)
U = runif(n, 0, 1)
#Membangkitkan status tersensor (0:censored 1:failure)
delta = rbinom(n,1,0.5) #tersensor 50%
delta10 = rbinom(n,1,0.9) #tersensor 10%
delta20 = rbinom(n,1,0.8) #tersensor 20%
delta30 = rbinom(n,1,0.7) #tersensor 30%
delta40 = rbinom(n,1,0.6) #tersensor 40%
delta60 = rbinom(n,1,0.4) #tersensor 60%
delta70 = rbinom(n,1,0.3) #tersensor 70%
delta80 = rbinom(n,1,0.2) #tersensor 80%
delta90 = rbinom(n,1,0.1) #tersensor 10%
# Menentukan koefisien variabel prediktor (beta)
b1 = 0.01 b2 = -0.015 b31 = -0.07 b32 = 0.04 b33 = -0.08 b41 = 0.015
b42 = 0.01 b43 = -0.03 b5 = -0.15 b6 = -0.08
beta = c (b1, b2, b31, b32, b33, b41, b42, b43, b5, b6)
#Membangkitkan Survival Time dengan pendekatan Distribusi Weibull
X.beta = cbind(X1, X2, X31, X32, X33, X41, X42, X43, X5, X6)%*%beta
lambda = 2
v = 2
T = (-1*log10(U)/(lambda*exp(X.beta)))^(1/v)
data= data.frame (T,delta,X1,X2,X31,X32,X33,X41,X42,X43,X5,X6,delta10,
                  delta20, delta30, delta40, delta60, delta70, delta80, delta90)
# Menyimpan Data (sample size=100, variabel prediktor=6 parametrik)
write.csv(data,file="E:/data100_6_parametrik.csv")
```

LAMPIRAN 15. Syntax R untuk Function yang Digunakan pada Model SURLS-SVM

```
#Function untuk menghitung nilai c-index
cindex = function(prog, Y, delta) {
  resh = function(data) {
    n = length(data)
    data = data.frame(no = 1:n, progx = data)
    b1 = c()
    for (i in 1:n) {
      b1 = rbind(b1,data)
    }
    b1 = as.matrix(b1[order(b1$no),]$progx)
    return(b1)
  }
  prog = (as.matrix(prog))
  n = nrow(prog)
  a = matrix(rep(prog,n),ncol=1)
  b = resh(prog)
  us = as.matrix(as.numeric((a-b)<0)+0.5*as.numeric((a-b)==0))
  ay = matrix(rep(Y,n),ncol=1)
  by = resh(Y)
  Ys = as.matrix(as.numeric((ay-by)<0))
  d1 = matrix(rep(delta,n), ncol = 1)
  d2 = resh(delta)
  comp = as.numeric(Ys==1 & d1==1)
  conc = comp*us*Ys
  c = sum(conc)/sum(comp)*100
  # hasil = data.frame(a=a,b=b,aminb=(a-b))
  return(c)
}

# Y dan delta untuk menyusun matriks D
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
```

LAMPIRAN 15. Syntax R untuk Function yang Digunakan pada Model SURLS-SVM
(Lanjutan)

```
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)
}

#Function untuk Kernel RBF
kernel_rbf = function(dataX,kernel_pars,dataXUrutY) {
  X = dataX
  Xt = dataXUrutY
  kpar = kernel_pars
  nr = dim(X)[1]
  nc = dim(X)[2]
  XXh1 = matrix(rep(apply(Xt*Xt,1,sum),nr),ncol = nr)
  XXh2 = matrix(rep(apply(X*X,1,sum),nr),ncol = nr)
  omega = XXh1+t(XXh2) - 2*(Xt%*%t(X))
  omega = exp(-omega/(2*kpar))
  return(omega)
}
```

LAMPIRAN 15. Syntax R untuk Function yang Digunakan pada Model SURLS-SVM
(Lanjutan)

```
#Function utama untuk model SURLSSVM
surlssvm = function(dataX, dataY, dataD, C, ker_par) {
  X    = dataX
  Y    = as.vector(dataY)
  delta = dataD
  C    = C
  ker_par = ker_par
  dimx  = dim(X)[1]
  si    = order(Y)
  Ybaru = Y[si]
  Xbaru = X[si,]
  delta = delta[si]
  # hitung kernel
  source("C:/Users/LENOVO/Documents/R/TESIS/rbf_kernel.R")
  K    = kernel_rbf(X, ker_par, Xbaru)
  K    = round(K,7)
  # hitung matriks D
  source("C:/Users/LENOVO/Documents/R/TESIS/makeDiffR.R")
  D    = makeDiff(Ybaru,delta)
  V    = matrix(1,dim(D)[1],1)
  V    = as.vector(V)
  ne   = dim(D)[1]
  A    = C*diag(V)%*%D%*%K%*%t(D)+diag(ne)
  ATA  = (A+t(A))/2
  ones = rep(1,ne)
  sol  = solve(ATA, C*ones)
  prog = t(sol%*%D%*%K)
  sol  = t(D)%*%sol
  sol[si] = sol
  z    = sum(sol)
  q    = sum(prog)
  prognew = prog[si]

  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  index = cindex(prognew, Ybaru, delta)

  list(alpha = sol, prog = prog, cindex = index)
}
```

LAMPIRAN 16. Syntax R untuk Analisis dengan Cox PHM dan SURLS-SVM serta menentukan Tuning Parameter Optimal dengan Metode *Grid Search*

```

Perf.coxphx = data.frame(C = c(), kern_par = c(), HR = c(), Logrank = c(),C.index=c())
Perf.surls  = data.frame(C = c(), kern_par = c(), HR = c(), Logrank = c(),C.index=c())
Perbandingan = data.frame(Data = c(), HR.COX = c(), HR.SURLS = c(), Logrank.COX
                          = c(), Logrank.SURLS = c(), C.index.COX = c(),
                          C.index.SURLS = c(),C = c(), kern_par = c())

n      = 1000
kern_par = c(2^(-4), 2^(-3), 2^(-2), 2^(-1), 2^(0), 2, 4) #Parameter Kernel
C = c(0.01,0.1,1)          #Parameter Model SURLSSVM
set.seed(1829)
for (i in 1){
  print(paste("data",i,"mulai"))
  data_file = paste0("E:/TESIS/DATASIMULASI17/Simulasi",i,".csv")
  data      = read.csv(data_file, sep = ",", header = TRUE)
  dataY     = data$T
  delta     = data$delta10          #berdasarkan nilai c-index terkecil
  dataX     =data.frame(data$X1,data$X2,data$X3,data$X4,data$X5,data$X61,data$X62,
data$X63,data$X71,data$X72,data$X73,data$X81,data$X82,data$X91,data$X92,data$
X101,data$X102,data$X11,data$X12,data$X13,data$X14,data$X15,data$X16,data$X17
)
  nameVar   = c("T","delta","X1","X2","X3","X4","X5","X61","X62","X63","X71",
"X72","X73","X81","X82","X91","X92","X101","X102","X11","X12","X13","X14"
,"X15","X16","X17")
  names(data) = nameVar
  library(survival)
  library(splines)
  print(paste("Estimasi COXPH"))
  coxphx    = coxph(Surv(dataY,delta)~X1+X2+X3+X4+X5+X61+X62+X63+X71+
X72+X73+X81+X82+X91+X92+X101+X102+X11+X12+X13+
X14+X15+X16+X17,data,method="efron")
  coxphx.b  = cbind(coxphx$coefficients)
  dataX     = as.matrix(dataX)
  ## PROGNOSTIC COX
  p.coxphx  = dataX %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)

#Menentukan grup berdasarkan model COX PHM
for (p in 1:n)
{
  if (p.coxphx[p]<p.coxphx.med) # 1:high risk
  {p.coxphx.hasil[p] = 1}
  else
  {p.coxphx.hasil[p] = 0} # 0 : low risk
}

```

LAMPIRAN 16. Syntax R untuk Analisis dengan Cox PHM dan SURLS-SVM serta menentukan Tuning Parameter Optimal dengan Metode *Grid Search* (Lanjutan)

```

for (ci in 1:length(C)) {
  for (ki in 1:length(kern_par)) {
    print(paste("Estimasi SURLS SVM dengan C =",C[ci],"dan Parameter Kernel = ",
kern_par[ki]))
    ## PROGNOSTIC SURLS-SVM
    source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
    surls      = surlssvm(dataX,dataY,delta,C[ci],kern_par[ki])
    p.surls    = surls$prog
    p.surls    = as.matrix(p.surls)
    p.surls.med = mean(p.surls)
    #Menentukan grup berdasarkan model SURLS-SVM
    for (m in 1:n)
    {
      if (p.surls[m]>p.surls.med) # 1:high risk
      {p.surls.hasil[m] = 1}
      else
      {p.surls.hasil[m] = 0} # 0 : low risk
    }
    Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
    g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
    g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
    d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
    for (j in 1:2){
      # HAZARD RATIO
      reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
      coeff.HR = reg.HR$coef
      HR = exp(coeff.HR)
      # LOG-RANK
      Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
      fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
      # Concordance index
      source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
      c_index = cindex(Prognostic[,j],dataY, delta)
      c.index = c_index
      if (j == 1) {
        Perf.coxphx.i = cbind(C = C[ci], kern_par = kern_par[ki], HR = HR, Logrank =
          Logrank$chisq, C.index = c.index)
      } else {
        Perf.surls.i = cbind(C = C[ci], kern_par = kern_par[ki],
          HR = HR, Logrank = Logrank$chisq,
          C.index = c.index) }
    }
  }
}

```

LAMPIRAN 16. Syntax R untuk Analisis dengan Cox PHM dan SURLS-SVM serta menentukan Tuning Parameter Optimal dengan Metode *Grid Search* (Lanjutan)

```
Perf.surls = rbind(Perf.surls, Perf.surls.i)

  Perbandinganx = data.frame(Data = i, HR.COX = Perf.coxphx.i[3], HR.SURLS =
Perf.surls.i[3],
                            Logrank.COX = Perf.coxphx.i[4], Logrank.SURLS = Perf.surls.i[4],
                            C.index.COX = Perf.coxphx.i[5], C.index.SURLS = Perf.surls.i[5],
                            C = C[ci], kern_par = kern_par[ki])
  print(Perbandinganx)
  Perbandingan = rbind(Perbandingan, Perbandinganx)
  write.csv(x = Perbandingan, file = "E:/TESIS/DATASIMULASI17/Perbandingan
Performace COXPH dan SURLS SVM Full Model Data1_0.1.csv")
}
}
}
```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM

```

Perf.coxphx = data.frame(Data = c(), Hapus.X = c(), HR = c(), Logrank = c(), C.index=c())
Perf.surls  = data.frame(Data = c(),Hapus.X = c(), HR = c(), Logrank = c(), C.index=c())
Perbandingan = data.frame(Data = c(), Hapus.X = c(), HR.COX = c(), HR.SURLS = c(),
                          Logrank.COX = c(), Logrank.SURLS = c(),C.index.COX = c(),
                          C.index.SURLS = c())
n          = 1000 #Jumlah sample size
C = 0.01          #parameter SURLSSVM
kern_par = 1     #parameter kernel
set.seed(1829)
#Feature selection untuk data simulasi 1 (X=17, n=1000)
for (i in 1){
  print(paste("data",i,"mulai"))
  data_file = paste0("E:/TESIS/DATASIMULASI17/Simulasi",i,".csv")
  data      = read.csv(data_file, sep = ",", header = TRUE)
  dataY     = data$T
  delta    = data$delta10
  dataX    = data[,-c(1,2,27,28,29,30,31,32,33,34)]
  nameVar  = c("T","delta","X1","X2","X3","X4","X5","X61","X62","X63","X71",
              "X72","X73","X81","X82","X91","X92","X101","X102","X11","X12",
              "X13","X14","X15","X16","X17")
  names(data) = nameVar
  library(survival)
  library(splines)
  #Feature Selection untuk X1 sampai X5
  for (xxx in 1:5) {
    namexxx = nameVar[-c(1,2,(xxx+2))]
    print(paste0("Estimasi COXPH X",xxx, " pada data ke-",i))
    dataxxx = dataX[,-xxx]
    names(dataxxx) = namexxx
    coxphx  = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
    coxphx.b = cbind(coxphx$coefficients)
    dataxxx  = as.matrix(dataxxx)
    rownames(coxphx.b) = namexxx
  ## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
  #Menentukan grup berdasarkan model COX PHM
  for (p in 1:n) {
    if (p.coxphx[p]<p.coxphx.med) # 1:high risk
      {p.coxphx.hasil[p] = 1}
    else
      {p.coxphx.hasil[p] = 0} # 0 : low risk
  }
}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```
print(paste("Estimasi SURLS SVM"))
print(paste("Estimasi SURLS SVM dengan C =",C,"dan Parameter Kernel = ",
           kern_par))
## PROGNOSTIC SURLS-SVM
source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
p.surls = surls$prog
p.surls = as.matrix(p.surls)
p.surls.med = mean(p.surls)
#Menentukan grup berdasarkan model SURLS-SVM
for (m in 1:n)
{
  if (p.surls[m]>p.surls.med) # 1:high risk
  {p.surls.hasil[m] = 1}
  else
  {p.surls.hasil[m] = 0} # 0 : low risk
}
Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR = exp(coeff.HR)
  # LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
  # Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
if (j == 1) {
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx, HR = HR, Logrank = Logrank$chisq,
                       C.index = c.index)
} else {
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx, HR = HR, Logrank = Logrank$chisq,
                      C.index = c.index)
}
}
}
```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx, HR.COX = Perf.coxphx.i[3],
                           HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],
                           Logrank.SURLS = Perf.surls.i[4], C.index.COX = Perf.coxphx.i[5],
                           C.index.SURLS = Perf.surls.i[5])

print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/DATASIMULASI17/Perbandingan
  Performace COXPH dan SURLS SVM Feature X1X5_data1.csv")
}
}
#Feature Selection untuk X6
for (xxx in 6) {
  namexxx = nameVar[-c(1,2,8,9,10)]
  print(paste0("Estimasi COXPH X",xxx, " pada data ke-",i))
  dataxxx = dataX[-c(6,7,8)]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
  for (p in 1:n)
  {
    if (p.coxphx[p]<p.coxphx.med) # 1:high risk
    {p.coxphx.hasil[p] = 1}
    else
    {p.coxphx.hasil[p] = 0} # 0 : low risk
  }
}
print(paste("Estimasi SURLS SVM"))
## PROGNOSTIC SURLS-SVM
source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
p.surls = surls$prog
p.surls = as.matrix(p.surls)
p.surls.med = mean(p.surls)

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

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
}
Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR = exp(coeff.HR)
  # LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
  # Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
if (j == 1) {
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx, HR = HR, Logrank = Logrank$chisq,
  C.index = c.index)
} else {
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx, HR = HR, Logrank = Logrank$chisq,
  C.index = c.index)
}
}
Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx, HR.COX = Perf.coxphx.i[3],
  HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],
  Logrank.SURLS = Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5],
  C.index.SURLS = Perf.surls.i[5])
print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPH dan
  SURLS SVM Feature X6_data1.csv")
}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```
#Feature selection untuk X7
for (xxx in 9) {
  namexxx = nameVar[-c(1,2,11,12,13)]
  print(paste0("Estimasi COXPH X",xxx-2, " pada data ke-",i))
  dataxxx = dataX[-c(9,10,11)]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
  ## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
  for (p in 1:n)
  {
    if (p.coxphx[p]<p.coxphx.med) # 1:high risk
    {p.coxphx.hasil[p] = 1}
    else
    {p.coxphx.hasil[p] = 0} # 0 : low risk
  }
}
print(paste("Estimasi SURLS SVM"))
## PROGNOSTIC SURLS-SVM
source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
p.surls = surls$prog
p.surls = as.matrix(p.surls)
p.surls.med = mean(p.surls)
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
}
}
Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR = exp(coeff.HR)
}
```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

# LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
# Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
  if (j == 1) {
Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx-2, HR = HR, Logrank = Logrank$chisq,
                      C.index = c.index)
  } else {
Perf.surls.i = cbind(Data = i, Hapus.X = xxx-2, HR = HR, Logrank = Logrank$chisq,
                    C.index = c.index)
  }
}
Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx-2," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx-2, HR.COX = Perf.coxphx.i[3],
                           HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],
                           Logrank.SURLS = Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5],
                           C.index.SURLS = Perf.surls.i[5])

print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPH dan
SURLS SVM Feature X7_data1.csv")
}
#Feature selection untuk variabel X8
for (xxx in 12) {
  namexxx = nameVar[-c(1,2,14,15)]
  print(paste0("Estimasi COXPH X",xxx-4, " pada data ke-",i))
  dataxxx = dataX[-c(12,13)]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

for (p in 1:n)
{
  if (p.coxphx[p]<p.coxphx.med) # 1:high risk
  {p.coxphx.hasil[p] = 1}
  else
  {p.coxphx.hasil[p] = 0} # 0 : low risk
}
print(paste("Estimasi SURLS SVM"))
## PROGNOSTIC SURLS-SVM
source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
surls      = surlssvm(dataxxx,dataY,delta,C,kern_par)
p.surls    = surls$prog
p.surls    = as.matrix(p.surls)
p.surls.med = mean(p.surls)
for (m in 1:n)
{
  if (p.surls[m]>p.surls.med) # 1:high risk
  {p.surls.hasil[m] = 1}
  else
  {p.surls.hasil[m] = 0} # 0 : low risk
}
Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR      = exp(coeff.HR)
  # LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit     = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
  # Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
if (j == 1) {
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx-4, HR = HR, Logrank = Logrank$chisq,
                        C.index = c.index)
  } else {
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx-4, HR = HR, Logrank = Logrank$chisq,
                      C.index = c.index) }}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx-4," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx-4, HR.COX = Perf.coxphx.i[3],
HR.SURLS = Perf.surls.i[3],Logrank.COX = Perf.coxphx.i[4], Logrank.SURLS =
Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5], C.index.SURLS = Perf.surls.i[5])
print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPH dan
SURLS SVM Feature X8_data1.csv")
}
#Feature selection untuk variabel X9
for (xxx in 14) {
  namexxx = nameVar[-c(1,2,16,17)]
  print(paste0("Estimasi COXPH X",xxx-5, " pada data ke-",i))
  dataxxx = dataX[-c(14,15)]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
  for (p in 1:n)
  {
    if (p.coxphx[p]<p.coxphx.med) # 1:high risk
    {p.coxphx.hasil[p] = 1}
    else
    {p.coxphx.hasil[p] = 0} # 0 : low risk
  }
  print(paste("Estimasi SURLS SVM"))
## PROGNOSTIC SURLS-SVM
  source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
  surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
  p.surls = surls$prog
  p.surls = as.matrix(p.surls)
  p.surls.med = mean(p.surls)

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

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
}
Prognostic = data.frame(COXPX = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPX = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR = exp(coeff.HR)
  # LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
  # Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
  if (j == 1) {
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx-5, HR = HR, Logrank = Logrank$Schisq,
    C.index = c.index)
  } else {
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx-5, HR = HR, Logrank = Logrank$Schisq,
    C.index = c.index)
  }
}
Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx-5," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx-5, HR.COX = Perf.coxphx.i[3],
  HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],
  Logrank.SURLS = Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5],
  C.index.SURLS = Perf.surls.i[5])
print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPX dan
  SURLS SVM Feature X9_data1.csv")
}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```
#Feature selection untuk variabel X10
for (xxx in 16) {
  namexxx = nameVar[-c(1,2,18,19)]
  print(paste0("Estimasi COXPH X",xxx-6, " pada data ke-",i))
  dataxxx = dataX[-c(16,17)]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
#PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
  for (p in 1:n)
  {
    if (p.coxphx[p]<p.coxphx.med) # 1:high risk
    {p.coxphx.hasil[p] = 1}
    else
    {p.coxphx.hasil[p] = 0} # 0 : low risk
  }
  print(paste("Estimasi SURLS SVM"))
#PROGNOSTIC SURLS-SVM
  source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
  surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
  p.surls = surls$prog
  p.surls = as.matrix(p.surls)
  p.surls.med = mean(p.surls)
  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
  }
  Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
  g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
  g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
  d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
  for (j in 1:2){
    # HAZARD RATIO
    reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
    coeff.HR = reg.HR$coef
    HR = exp(coeff.HR)
```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```

# LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
# Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
if (j == 1) {
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx-6, HR = HR, Logrank = Logrank$chisq,
                        C.index = c.index)
  } else {
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx-6, HR = HR, Logrank = Logrank$chisq,
                      C.index = c.index)
  }
}
Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)
Perf.surls = rbind(Perf.surls, Perf.surls.i)
print(paste0("data ",i," - menghapus X",xxx-6," selesai"))
Perbandinganx = data.frame(Data = i, Hapus.X = xxx-6, HR.COX = Perf.coxphx.i[3],
                           HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],
                           Logrank.SURLS = Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5],
                           C.index.SURLS = Perf.surls.i[5])

print(Perbandinganx)
Perbandingan = rbind(Perbandingan, Perbandinganx)
write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPH dan
SURLS SVM Feature X10_data1.csv")
}
#Feature selection untuk variabel X11 sampai X17
for (xxx in 18:24) {
  namexxx = nameVar[-c(1,2,(xxx-7))]
  print(paste0("Estimasi COXPH X",xxx-7, " pada data ke-",i))
  dataxxx = dataX[,-xxx]
  names(dataxxx) = namexxx
  coxphx = coxph(Surv(dataY,delta)~as.matrix(dataxxx),data,method="efron")
  coxphx.b = cbind(coxphx$coefficients)
  dataxxx = as.matrix(dataxxx)
  rownames(coxphx.b) = namexxx
  ## PROGNOSTIC COX
  p.coxphx = dataxxx %*% coxphx.b
  p.coxphx.med = mean(p.coxphx)
}

```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```
#Determined group by COX PHM
for (p in 1:n)
{
  if (p.coxphx[p]<p.coxphx.med) # 1:high risk
  {p.coxphx.hasil[p] = 1}
  else
  {p.coxphx.hasil[p] = 0} # 0 : low risk
}
print(paste("Estimasi SURLS SVM"))
## PROGNOSTIC SURLS-SVM
source("C:/Users/LENOVO/Documents/R/TESIS/surlssvm.R")
surls = surlssvm(dataxxx,dataY,delta,C,kern_par)
p.surls = surls$prog
p.surls = as.matrix(p.surls)
p.surls.med = mean(p.surls)
for (m in 1:n)
{
  if (p.surls[m]>p.surls.med) # 1:high risk
  {p.surls.hasil[m] = 1}
  else
  {p.surls.hasil[m] = 0} # 0 : low risk
}
Prognostic = data.frame(COXPH = p.coxphx, SURLSSVM = p.surls)
g.coxphx = as.matrix(p.coxphx.hasil[1:n]) #prognostic Cox
g.surls = as.matrix(p.surls.hasil[1:n]) #prognostic SURLS-SVM
d.group = cbind(COXPH = g.coxphx, SURLSSVM = g.surls)
for (j in 1:2){
  # HAZARD RATIO
  reg.HR = coxph(Surv(dataY,delta)~d.group[,j],method="efron")
  coeff.HR = reg.HR$coef
  HR = exp(coeff.HR)
  # LOG-RANK
  Logrank = survdiff(Surv(dataY,delta)~d.group[,j])
  fit = survfit(Surv(ceiling(dataY),delta)~d.group[,j])
  # Concordance index
  source("C:/Users/LENOVO/Documents/R/TESIS/cindex.R")
  c_index = cindex(Prognostic[,j],dataY, delta)
  c.index = c_index
}
```

LAMPIRAN 17. Syntax R untuk *Feature Selection* dengan Metode *Backward* pada model Cox PHM dan SURLS-SVM (lanjutan)

```
if (j == 1) {  
  Perf.coxphx.i = cbind(Data = i, Hapus.X = xxx-7, HR = HR, Logrank = Logrank$chisq,  
                        C.index = c.index)  
  } else {  
  Perf.surls.i = cbind(Data = i, Hapus.X = xxx-7, HR = HR, Logrank = Logrank$chisq,  
                       C.index = c.index)  
  }  
  }  
  Perf.coxphx = rbind(Perf.coxphx, Perf.coxphx.i)  
  Perf.surls = rbind(Perf.surls, Perf.surls.i)  
  print(paste0("data ",i," - menghapus X",xxx-7," selesai"))  
  Perbandinganx = data.frame(Data = i, Hapus.X = xxx-7, HR.COX = Perf.coxphx.i[3],  
                              HR.SURLS = Perf.surls.i[3], Logrank.COX = Perf.coxphx.i[4],  
                              Logrank.SURLS = Perf.surls.i[4],C.index.COX = Perf.coxphx.i[5],  
                              C.index.SURLS = Perf.surls.i[5])  
  print(Perbandinganx)  
  Perbandingan = rbind(Perbandingan, Perbandinganx)  
  write.csv(x = Perbandingan, file = "E:/TESIS/Perbandingan Performace COXPH dan  
SURLS SVM Feature X1117_data1.csv")  
}
```

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini, mahasiswa Departemen Statistika FMKSD ITS:

Nama : Halwa Annisa Khoiri

NRP : 06211550010201

Program Studi : Magister Statistika / ~~Doktor Ilmu Statistik~~

menyatakan bahwa data yang digunakan dalam Tesis / ~~Disertasi~~ ini merupakan data primer yang diambil dari:

Sumber : Rekam medik RSUD dr. Soetomo Surabaya

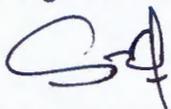
Keterangan : Data yang diambil adalah data pasien kanker serviks yang menjalani rawat inap di RSUD dr. Soetomo Surabaya tahun 2013-2016

Surat pernyataan ini dibuat dengan sebenarnya. Apabila terdapat pemalsuan data maka saya siap menerima sanksi sesuai aturan yang berlaku.

Surabaya, 17 Januari 2018

Mengetahui

Pembimbing Tesis/~~Disertasi~~



(Santi Wulan Purnami, M.Si, Ph.D)
NIP. 19720923 199803 2 001

Mahasiswa



(Halwa Annisa Khoiri)
NRP. 06211550010201

*(coret yang tidak perlu)

** Mohon dituliskan data primer apabila datanya dikumpulkan sendiri

BIOGRAFI PENULIS



Penulis lahir di Kota Madiun Jawa Timur pada tanggal 6 Juli 1992 dengan nama lengkap Halwa Annisa Khoiri anak pertama dari dua bersaudara, putri pasangan Sunaryo dan Wahyuni Aminah. Penulis menempuh pendidikan formal di RA Islamiyah Rejomulyo Madiun (1997-1998), MI Islamiyah Rejomulyo Madiun (1998-2004), SMP Negeri 8 Madiun (2004-2007) dan SMA Negeri 5 Madiun (2007-2010). Penulis melanjutkan pendidikan ke jenjang S1 di Jurusan Matematika ITS (2010-2014). Penulis berkesempatan melanjutkan studi ke jenjang S2 di Program Pascasarjana Statistika FMIPA ITS (2015-2017) dengan Beasiswa dari Lembaga Pengelola Dana Pendidikan (LPDP).

Saran, kritik, dan pertanyaan mengenai tesis ini dapat disampaikan ke alamat email halwa.nisa92@gmail.com.

(halaman ini sengaja dikosongkan)