



TUGAS AKHIR - SS141501

**KLASIFIKASI HASIL *PAP SMEAR TEST* SEBAGAI
UPAYA PENCEGAHAN SEKUNDER PENYAKIT
KANKER SERVIKS DI RUMAH SAKIT “X”
SURABAYA MENGGUNAKAN *PIECEWISE
POLYNOMIAL SMOOTH SUPPORT VECTOR
MACHINE (PPSSVM)***

Mukti Ratna Dewi
NRP 1311 100 081

Dosen Pembimbing
Santi Wulan Purnami, S.Si, M.Si, Ph.D

JURUSAN STATISTIKA
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT- SS141501

**CLASSIFICATION OF PAP SMEAR TEST RESULT
AS SECONDARY PREVENTION EFFORT IN
CERVICAL CANCER AT “X” HOSPITAL SURABAYA
USING PIECEWISE POLYNOMIAL SMOOTH
SUPPORT VECTOR MACHINE (PPSSVM)**

Mukti Ratna Dewi
NRP 1311 100 081

Supervisor
Santi Wulan Purnami, S.Si, M.Si, Ph.D

DEPARTMENT Of STATISTICS
Faculty of Mathematics and Natural Science
Sepuluh Nopember Institute Of Technology
Surabaya 2015

LEMBAR PENGESAHAN

**KLASIFIKASI HASIL *PAP SMEAR TEST* SEBAGAI
UPAYA PENCEGAHAN SEKUNDER PENYAKIT
KANKER SERVIKS DI RUMAH SAKIT "X" SURABAYA
MENGUNAKAN *PIECEWISE POLYNOMIAL SMOOTH
SUPPORT VECTOR MACHINE (PPSSVM)***

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

MUKTI RATNA DEWI
NRP 1311 100 081

Disetujui oleh Pembimbing Tugas Akhir
Santi Wulan Purnami, S.Si, M.Si, Ph.D
NIP : 19720923 199803 2 001

Mengetahui

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Jurusan Statistika FMIPA-ITS



Dr. Muhammad Mashuri, MT
NIP 19620408 198701 1 001

SURABAYA, JANUARI 2015

Klasifikasi Hasil *Pap Smear Test* Sebagai Upaya Pencegahan Sekunder Penyakit Kanker Serviks di Rumah Sakit “X” Surabaya Menggunakan *Piecewise Polynomial Smooth Support Vector Machine* (PPSSVM)

Nama Mahasiswa : Mukti Ratna Dewi
NRP : 1311 100 081
Jurusan : Statistika
Pembimbing : Santi Wulan Purnami, S.Si, M.Si, Ph.D

Abstrak

Kanker serviks merupakan kanker yang menyerang leher rahim dan ditandai dengan pertumbuhan abnormal pada sel leher rahim. Berdasarkan data Riset Kesehatan Dasar tahun 2013, kanker serviks merupakan salah satu kanker yang paling sering menyerang perempuan di Indonesia. Oleh karena itu pencegahan sekunder melalui deteksi dini penting dilakukan. Penelitian ini mengangkat permasalahan klasifikasi hasil pap smear test di Rumah Sakit “X” Surabaya tahun 2010 menggunakan metode PPSSVM yang didasarkan pada ketujuh faktor risiko eksternal, yaitu usia, usia pertama kali menstruasi, usia pertama kali melahirkan, siklus menstruasi, penggunaan alat kontrasepsi, paritas, dan riwayat keguguran. Hasil penelitian menunjukkan bahwa dari 3586 pasien yang melakukan pap smear test, 33% di antaranya menunjukkan hasil abnormal. Pada hasil klasifikasi pap smear test, metode PPSSVM1 memiliki rata-rata tingkat sensitivitas tertinggi, yaitu 91,22% pada data training dan 94,66% pada data testing. Sementara metode PPSSVM2 memiliki rata-rata nilai akurasi dan spesifisitas tertinggi, yaitu 89,22% dan 88,53% pada data training serta 92,84% dan 91,12% untuk data testing. Berdasarkan kurva ROC dan nilai AUC, kedua metode PPSSVM sama-sama baik dalam mengklasifikasikan hasil pap smear test di Rumah Sakit “X” Surabaya. Model PPSSVM1 lebih baik digunakan bila tingkat sensitivitas lebih diutamakan. Sementara bila tingkat spesifisitas lebih diutamakan maka model PPSSVM2 lebih baik untuk digunakan.

Kata Kunci : *kanker serviks, pap smear test, klasifikasi, SSVM, PPSSVM*

Classification Of Pap Smear Test Result As Secondary Prevention Effort in Cervical Cancer at “X” Hospital Surabaya Using Piecewise Polynomial Smooth Support Vector Machine (PPSSVM)

Name : Mukti Ratna Dewi
NRP : 1311 100 081
Department : Statistika
Supervisor : Santi Wulan Purnami, S.Si, M.Si, Ph.D

Abstract

Cervical cancer is a cancer arising from the cervix and is due to the abnormal growth of cells in the cervix. Based on data from the Health Research in 2013, in Indonesia cervical cancer is one of the most common cancer in women. Therefore, secondary prevention through early detection is important. This study raises the issue of classification of pap smear test in the "X" Hospital Surabaya in 2010 using PPSSVM method which is based on the seven external risk factors, namely age, age at menarche, age at first childbirth, menstrual period, contraceptive usage, parity, and miscarriage history. The results showed that among 3000 patients undergoing Pap smears, 33% of them showed abnormal results. In the classification results of pap smear test, PPSSVM1 method has the highest average level of sensitivity, which is 91,22% at training data and 94,66% at testing data. While PPSSVM2 method has the highest average value of accuracy and specificity level, ie 89,22% and 88,53% on the training data and 92,84% and 91,12% for testing data. Based on the ROC curve and AUC values, both PPSSVM methods are equally excellent at classifying Pap test results in the "X" Hospital Surabaya. PPSSVM1 is better used if level of sensitivity is preferred. Meanwhile, when the level of specificity is preferred then PPSSVM2 is better to be used.

Keywords : cervical cancer, pap smear test, classification, SSVM, PPSSVM

KATA PENGANTAR

Alhamdulillah.... Puji syukur kepada Allah SWT Yang Maha Esa dan Maha Kuasa atas segala limpahan anugerah, rahmat, dan pertolongan sehingga penulis dapat menyelesaikan Tugas Akhir dengan judul

KLASIFIKASI HASIL *PAP SMEAR TEST* SEBAGAI UPAYA PENCEGAHAN SEKUNDER PENYAKIT KANKER SERVIKS DI RUMAH SAKIT “X” SURABAYA MENGGUNAKAN *PIECEWISE POLYNOMIAL SMOOTH SUPPORT VECTOR MACHINE* (PPSSVM)

Shalawat serta salam semoga senantiasa tercurahkan kepada Nabi Muhammad SAW beserta para keluarga dan sahabat yang senantiasa membantu perjuangan beliau dalam menegakkan Dinullah di muka bumi ini.

Keberhasilan penyusunan Tugas Akhir ini tidak terlepas dari peran banyak pihak yang memberikan bantuan baik moril maupun material. Oleh karena itu penulis ingin menyampaikan ucapan terimakasih yang tiada hingganya kepada.

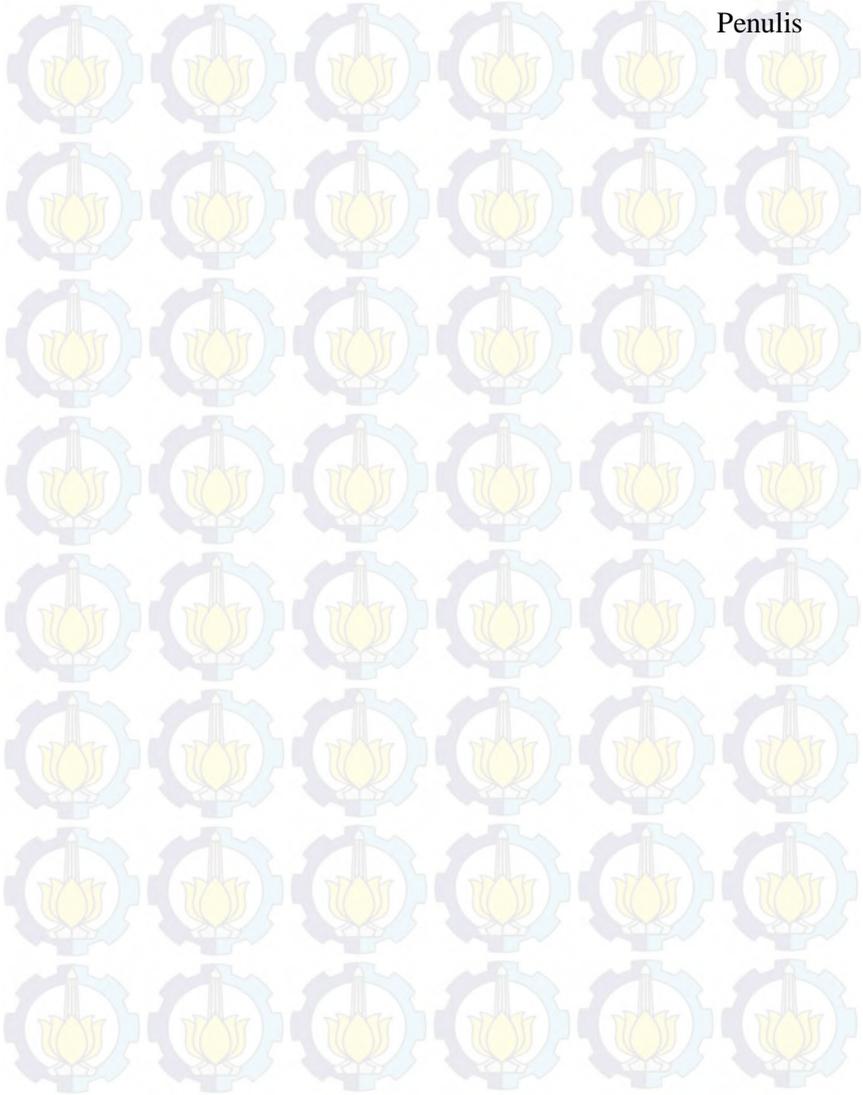
1. Ibu dan bapak yang menjadi sumber semangat di setiap kesempatan, dua orang terkasih yang menjadi *goal* di setiap cita-citaku. Terimakasih atas segala pengorbanan, doa, senyuman, kesabaran, dan semangat tiada henti yang kalian berikan kepada anakmu ini.
2. Ibu Santi Wulan Purnami, S.Si, M.Si, Ph.D selaku dosen pembimbing yang telah begitu banyak meluangkan waktu untuk membimbing dan mengarahkan penulis selama menyusun Tugas Akhir ini.
3. Ibu Ir. Sri Pingit Wulandari, M.Si dan Bapak Dr. Bambang Widjanarko Otok, M.Si selaku dosen penguji yang telah memberikan kritik dan saran.

4. Bapak Dr. Muhammad Mashuri, MT selaku Ketua Jurusan Statistika, Ibu Dra. Lucia Aridinanti, MT selaku Ketua Prodi S1 Jurusan Statistika, dan Bapak Dr. Sony Sunaryo, M.Si selaku dosen wali.
5. Semua dosen dan karyawan Statistika ITS atas segala ilmu dan *support* yang diberikan selama penulis kuliah. Pak Yusuf, terimakasih banyak atas kesabarannya dalam memberikan ijin untuk menggunakan komputer di laboratorium yang *nginep* sampai berhari-hari. Kalau *ndak* diijinkan mungkin tugas akhir ini *nggak* bisa selesai, pak. Hehehe...
6. Semua teman-teman Statistika ITS, *especially* Pejuang PW ITS 111 yang sama-sama merasakan pahit manis kehidupan mahasiswa akhir. Wahendra, *thanks a lot ya dude* karena berkat kamu aku jadi bisa *join* dengan Bu Santi dan dapat topik Tugas Akhir jauh-jauh hari sebelum semester 7 dimulai. Epha juga, *thanks bro* atas modifan *syntax*-nya. *It was a big help for me!*
7. *Arek-arek kos sing koplak, thanks* untuk semua canda tawa dan hiburan yang diberikan. Kalian akan menjadi salah satu yang paling aku rindukan selepas lulus nanti *rek!*
8. Keluarga besar tercinta atas semangat dan doa yang diberikan. *Mpok Luci* yang dengan sabar menemani kegalauanku, *thanks mpok* atas berbagai saran dan dorongan yang *mpok* berikan selama ini. Lihatlah, secepatnya aku akan menyalip kesuksesanmu *mpok!*
9. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang juga telah memberikan dukungan selama proses penyelesaian Tugas Akhir ini.

Penulis menyadari dalam penulisan Tugas Akhir ini masih jauh dari kesempurnaan. Oleh karena itu, saran dan masukan dari semua pihak sangat penulis harapkan demi perbaikan dan kesempurnaan. Penulis berharap semoga Tugas Akhir ini dapat memberikan manfaat dan sumbangsih bagi perkembangan ilmu pengetahuan.

Surabaya, Januari 2015

Penulis



DAFTAR ISI

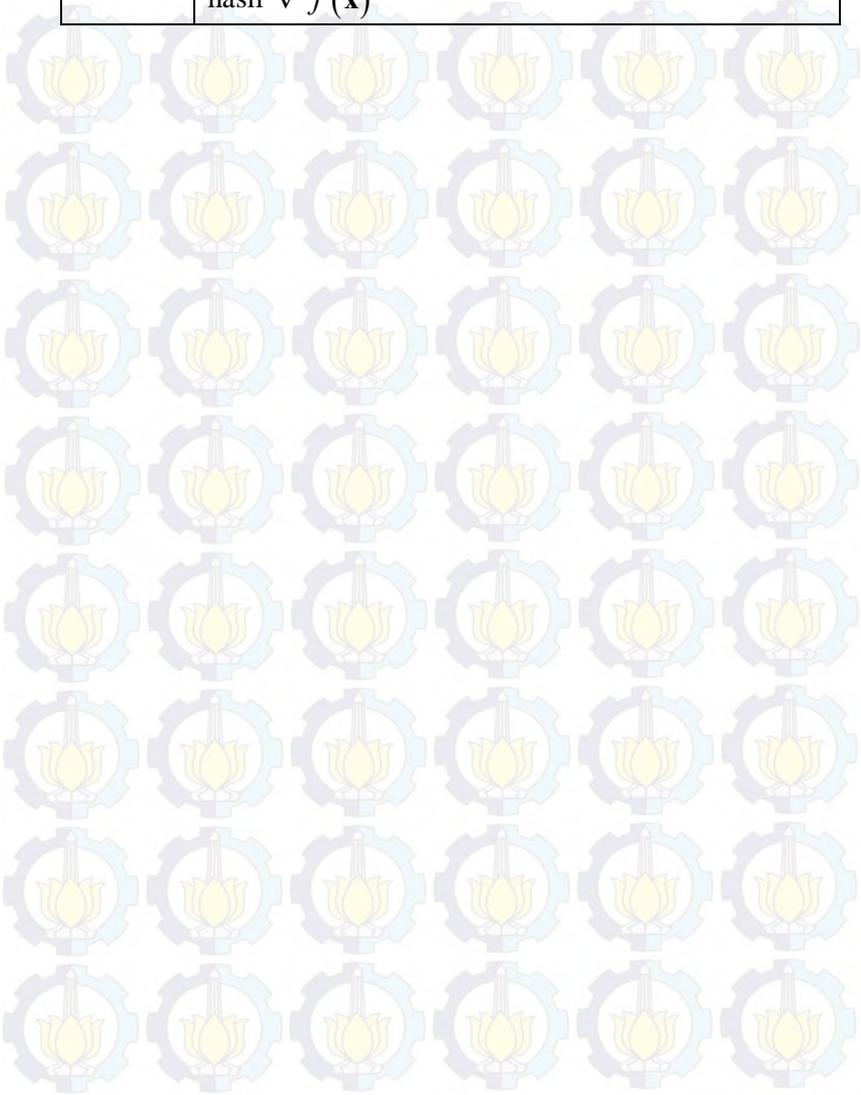
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
ABSTRAK	iv
KATA PENGANTAR	viii
DAFTAR ISI	xii
DAFTAR NOTASI	xiv
DAFTAR TABEL	xvi
DAFTAR GAMBAR	xviii
DAFTAR LAMPIRAN	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 <i>Support Vector Machine (SVM)</i>	7
2.1.1 <i>Linearly Separable Case</i>	7
2.1.2 <i>Linearly Nonseparable Case</i>	9
2.1.3 <i>Nonlinear Case</i>	11
2.2 <i>Smooth Support Vector Machine (SSVM)</i>	14
2.2.1 <i>SSVM Linier</i>	14
2.2.2 <i>SSVM Nonlinier</i>	18
2.2.3 <i>Contoh Implementasi Algoritma Newton Armidjo</i> <i>Pada SSVM</i>	19
2.3 <i>Polynomial Smoothing Function</i>	26
2.4 <i>Seleksi Paramater</i>	28
2.5 <i>Evaluasi Performasi Klasifikasi</i>	30
2.5.1 <i>Cross Validation</i>	31
2.5.2 <i>Kurva Receiving Operating Characteristic (ROC)</i>	32
2.6 <i>Kanker Serviks</i>	37
2.6.1 <i>Faktor Risiko</i>	38

2.6.2	<i>Pap Smear Test</i>	41
2.7	Penelitian Sebelumnya.....	41
BAB III METODOLOGI PENELITIAN		43
3.1	Sumber Data.....	43
3.2	Variabel Penelitian	43
3.3	Metode Analisis	44
BAB IV ANALISIS DAN PEMBAHASAN		49
4.1	Karakteristik Pasien Berdasarkan Faktor-Faktor Risiko Penyebab Kanker Serviks	49
4.1.1	Karakteristik Pasien Berdasarkan Faktor Usia, Usia Pertama Kali Menstruasi, dan Usia Pertama Kali Melahirkan.....	49
4.1.2	Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi.....	52
4.1.3	Karakteristik Pasien Berdasarkan Faktor Paritas.....	54
4.1.4	Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi	56
4.1.5	Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran.....	58
4.2	Klasifikasi Hasil <i>Pap Smear Test</i> Menggunakan Metode PPSSVM.....	60
4.2.1	Seleksi Parameter	61
4.2.2	Pembentukan Model Klasifikasi Hasil <i>Pap Smear Test</i>	62
4.2.3	Evaluasi Klasifikasi Hasil <i>Pap Smear Test</i>	63
4.2.4	Pemilihan Model Terbaik untuk Klasifikasi Hasil <i>Pap Smear</i>	70
BAB V KESIMPULAN DAN SARAN		81
5.1	Kesimpulan	81
5.2	Saran.....	82
DAFTAR PUSTAKA		83
LAMPIRAN		89

DAFTAR NOTASI

$\mathbf{x}^T \mathbf{y}$	Hasil skalar (<i>inner product</i>) dari dua vektor kolom \mathbf{x} dan \mathbf{y} pada dimensi real n
x_+	<i>plus function</i> , didefinisikan sebagai $(x_+)_i = \max\{0, x_i\}$, $i = 1, 2, \dots, n$
\mathbf{A}	matriks bilangan <i>real</i> berukuran $m \times n$
\mathbf{A}_i	baris ke- i dari matriks \mathbf{A} yang merupakan sebuah vektor baris dalam R^n
\mathbf{A}_j	kolom ke- j dari matriks \mathbf{A}
\mathbf{e}	vektor satu pada ruang dimensi real tertentu
$\mathbf{0}$	vektor nol pada ruang dimensi real tertentu
\mathbf{I}	matriks identitas di ruang dimensi real tertentu
$\text{sign}(x)$	<i>sign function</i> yang didefinisikan sebagai $\text{sign}(x) = \begin{cases} -1 & \text{jika } x < 0 \\ 0 & \text{jika } x = 0 \\ 1 & \text{jika } x > 0 \end{cases}$
$\ x\ _p$	p -norm untuk $\mathbf{x} \in R^n$ di mana $\ x\ _p = \left(\sum_{i=1}^n x_i ^p \right)^{\frac{1}{p}} \quad \text{dan} \quad \ x\ _p = \max_{1 \leq i \leq n} x_i $
$\ \bullet\ $	<i>general norm</i> pada R^n
$\ \bullet\ '$	<i>dual norm</i> pada R^n didefinisikan sebagai $\ x\ ' = \max_{\ y\ =1} \mathbf{x}^T \mathbf{y}$
f	fungsi bilangan <i>real</i> yang didefinisikan pada ruang <i>real</i> berdimensi n
$\nabla f(\mathbf{x})$	gradien dari f pada \mathbf{x} dan merupakan sebuah vektor baris pada R^n
$\nabla^2 f(\mathbf{x})$	turunan parsial kedua dari f terhadap \mathbf{x}

$\Psi_{n \times n}$	Matriks Hessian berukuran $n \times n$ yang merupakan hasil $\nabla^2 f(\mathbf{x})$
---------------------	--



DAFTAR TABEL

Tabel 2.1	Data Ilustrasi Newton-Armidjo	19
Tabel 2.2	Contoh Data <i>Training</i>	24
Tabel 2.3	Tabel Klasifikasi	30
Tabel 2.4	Kategori Pengklasifikasian Model Berdasarkan Nilai AUC.....	35
Tabel 2.5	Hasil Klasifikasi Biner	35
Tabel 2.6	Prediksi Kelas Berdasarkan Nilai <i>Cutoff</i>	36
Tabel 2.7	Koordinat Titik Kurva ROC	36
Tabel 3.1	Gambaran Sitologi Hasil <i>Pap Smear Test</i>	43
Tabel 3.2	Variabel Prediktor	44
Tabel 3.3	Struktur Data Penelitian	44
Tabel 4.1	Statistika Deskriptif Data Pasien <i>Pap Smear Test</i> ..	50
Tabel 4.2	Statistika Deskriptif Pasien Pada Kelompok Hasil <i>Pap Smear</i>	51
Tabel 4.3	Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi	52
Tabel 4.4	Karakteristik Pasien Berdasarkan Faktor Paritas	54
Tabel 4.5	Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi	56
Tabel 4.6	Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran	58
Tabel 4.7	Hasil dan Waktu Komputasi Seleksi Parameter	61
Tabel 4.8	Tabel Klasifikasi Metode SSVM	65
Tabel 4.9	Tabel Klasifikasi Metode PPSSVM1	66
Tabel 4.10	Tabel Klasifikasi Metode PPSSVM2.....	67
Tabel 4.11	Evaluasi Performasi Model SSVM, PPSSVM1, dan PPSSVM2 (%) Pada Data <i>Training</i>	68
Tabel 4.12	Evaluasi Performasi Model SSVM, PPSSVM1, dan PPSSVM2 (%) Pada Data <i>Testing</i>	69
Tabel 4.13	$s_1(c)$ dan $s_0(c)$ Pada Masing-Masing Fold Metode SSVM.....	72
Tabel 4.14	Koordinat Kurva ROC Masing-Masing <i>Fold</i> SSVM	72

Tabel 4.15	Nilai AUC Pada Masing-Masing <i>Fold</i> SSVM.....	72
Tabel 4.16	$s_1(c)$ dan $s_0(c)$ Pada Masing-Masing <i>Fold</i> Metode PPSSVM1	73
Tabel 4.17	Koordinat Kurva ROC Masing-Masing <i>Fold</i> PPSSVM1	74
Tabel 4.18	Nilai AUC Pada Masing-Masing <i>Fold</i> PPSSVM1 ..	75
Tabel 4.19	$s_1(c)$ dan $s_0(c)$ Pada Masing-Masing <i>Fold</i> Metode PPSSVM2	75
Tabel 4.20	Koordinat Kurva ROC Masing-Masing <i>Fold</i> PPSSVM2	75
Tabel 4.21	Nilai AUC Pada Masing-Masing <i>Fold</i> PPSSVM2 ..	75
Tabel 4.22	Luas AUC dari Model SSVM, PPSSVM1, dan PPSSVM2	77

DAFTAR GAMBAR

Gambar 2.1	Ilustrasi dari <i>Linearly Separable Case</i> Pada SVM	8
Gambar 2.2	Ilustrasi dari <i>Linearly Nonseparable Case</i> Pada SVM.....	9
Gambar 2.3	Ilustrasi dari <i>Nonlinear Case</i> pada SVM.....	12
Gambar 2.4	<i>Smooth Approximation</i> untuk <i>Plus Function</i>	15
Gambar 2.5	<i>Scatter Plot</i> Data Ilustrasi Newton Armidjo.....	19
Gambar 2.6	Empat Kurva ROC dengan Nilai AUC yang Berbeda	33
Gambar 2.7	Kurva ROC yang Terbentuk.....	37
Gambar 3.1	Tahapan Penelitian.....	46
Gambar 3.2	Algoritma Newton Armidjo.....	47
Gambar 4.1	Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi Pada Masing-Masing Hasil <i>Pap Smear Test</i>	52
Gambar 4.2	Karakteristik Pasien <i>Pap Smear Test</i> Berdasarkan Faktor Siklus Menstruasi.....	53
Gambar 4.3	Karakteristik Pasien Berdasarkan Faktor Paritas Pada Masing-Masing Hasil <i>Pap Smear Test</i>	54
Gambar 4.4	Karakteristik Pasien <i>Pap Smear Test</i> Berdasarkan Faktor Paritas.....	55
Gambar 4.5	Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi Pada Masing-Masing Hasil <i>Pap Smear Test</i>	57
Gambar 4.6	Karakteristik Pasien <i>Pap Smear Test</i> Berdasarkan Faktor Penggunaan Alat Kontrasepsi.....	57
Gambar 4.7	Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran Pada Masing-Masing Hasil <i>Pap Smear Test</i>	59
Gambar 4.8	Karakteristik Pasien <i>Pap Smear Test</i> Berdasarkan Faktor Riwayat Keguguran.....	59
Gambar 4.9	Perbandingan Kurva ROC dari Kelima <i>Fold</i> SSVM.....	73

Gambar 4.10 Perbandingan Kurva ROC dari Kelima <i>Fold</i> PPSSVM1.....	74
Gambar 4.11 Perbandingan Kurva ROC dari Kelima <i>Fold</i> PPSSVM2.....	76
Gambar 4.12 Kurva ROC untuk SSVM, PPSSVM1, dan PPSSVM2.....	77
Gambar 4.13 Kurva ROC untuk SSVM, PPSSVM1, dan PPSSVM2 (Diperbesar)	78

DAFTAR LAMPIRAN

Lampiran 1.	Data Hasil <i>Pap Smear Test</i>	89
Lampiran 2.	<i>Syntax</i> MATLAB untuk Seleksi Parameter	90
Lampiran 3.	<i>Syntax</i> MATLAB untuk Pembentukan Model SSVM, PPSSVM1, dan PPSSVM2	98
Lampiran 4.	<i>Syntax</i> MATLAB untuk Prediksi	102
Lampiran 5.	Parameter u dan γ untuk Model SSVM, PPSSVM1, dan PPSSVM2	103
Lampiran 6.	Penghitungan Akurasi, Sensitivitas, dan Spesifisitas	104
Lampiran 7.	<i>Output</i> SPSS Kurva ROC dan AUC	107

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saat ini penyakit tidak menular, termasuk kanker menjadi masalah kesehatan utama baik di dunia maupun di Indonesia. Kanker serviks adalah salah satu kanker yang paling sering diderita oleh perempuan di seluruh dunia dan menduduki peringkat keempat paling umum pada wanita dan secara keseluruhan ketujuh, dengan perkiraan 528.000 kasus baru pada tahun 2012 (Globocan, 2012). Setiap tahunnya, lebih dari 270.000 wanita meninggal karena kanker serviks; lebih dari 85% dari kematian ini berada di negara-negara berpenghasilan rendah dan menengah (WHO, 2013).

Di Indonesia, prevalensi penyakit kanker juga cukup tinggi. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2013, kanker yang paling sering menyerang perempuan adalah kanker payudara dan kanker serviks dengan 5.349 kasus (12,8%) untuk kanker serviks (Dinas Kesehatan, 2014). Hal ini menunjukkan bahwa kanker serviks adalah penyakit yang sangat berbahaya di Indonesia. Berdasarkan fakta tersebut dibutuhkan suatu langkah strategis untuk deteksi dini kanker serviks di Indonesia.

Penyakit kanker serviks belum diketahui penyebabnya secara pasti, namun berdasarkan penelitian yang telah berkembang didapatkan beberapa faktor risiko yang dapat memicu munculnya penyakit kanker serviks. Faktor risiko yang paling penting dari kanker serviks adalah infeksi *Human Papilloma Virus* (HPV) (IARC, 2014). Di samping itu, faktor usia (Nurwijaya, et al., 2010), usia saat pertama kali melahirkan (Cancer Research UK, 2014), usia saat menstruasi pertama kali (Yakasai, et al., 2013), jumlah anak (Fatmawati, 2012), penggunaan alat kontrasepsi (Sjamsudin, 2001; Dalimartha, 2004; Tira, 2008), dan riwayat keguguran juga merupakan beberapa faktor risiko lain dari kanker serviks (Abbas, 2003; Tira, 2008).

Kanker serviks sangat mungkin untuk dicegah dan sebuah program deteksi dini yang efektif dapat mengurangi angka kejadian dan kematian yang berhubungan dengan kanker ini. Terdapat dua jenis pencegahan untuk kanker serviks yaitu pencegahan primer dan sekunder. Pencegahan primer berkaitan dengan usaha untuk mencegah terjadinya infeksi HPV dan melakukan vaksin HPV, sementara pencegahan sekunder berkaitan dengan deteksi dini untuk kanker serviks (Nurwijaya, et al., 2010). Oleh karena itu, deteksi dini kanker serviks penting dilakukan untuk mencegah risiko lebih tinggi dari kanker serviks. Salah satu kegiatan deteksi dini kanker serviks yang paling umum di Indonesia adalah menggunakan metode *pap smear*. Pemeriksaan ini bermanfaat sebagai tes skrining awal untuk mendeteksi adanya kanker serviks. Suatu prosedur *pap smear* konvensional yang dilakukan dengan sangat baik dapat menghasilkan akurasi hasil sebesar 76% sampai dengan 94% (Nurwijaya, et al., 2010). Terdapat beberapa faktor yang menyebabkan hasil tes tidak akurat, antara lain sampel yang diambil kurang memadai, *human error* dari dokter atau perawat yang melakukan pemeriksaan, serta proses laboratorium yang tidak baik. Oleh karena itu diperlukan suatu penelitian untuk mendapatkan suatu model yang mampu mengklasifikasikan hasil *pap smear* dengan tingkat kesalahan klasifikasi yang minimal.

Beberapa metode klasifikasi hasil *Pap Smear Test* dikembangkan dari tahun ke tahun. Penelitian menggunakan berbagai metode statistik untuk klasifikasi deteksi dini kanker serviks juga telah banyak dilakukan. Tahun 2011, Yastuti menggunakan metode *Bagging Multivariate Adaptive Regression Splines* (MARS) untuk klasifikasi hasil *pap smear* kanker serviks di Rumah Sakit “X” Surabaya dan mendapatkan ketepatan klasifikasi sebesar 79,67% (Yastuti, 2011). Tahun berikutnya, Intansari melakukan penelitian dengan kasus yang sama menggunakan metode *logistic regression* dengan ketepatan klasifikasi sebesar 66,7% yang kemudian meningkat menjadi

70,74% setelah menggunakan pendekatan *bagging* regresi logistik (Intansari, 2012).

Salah satu metode klasifikasi yang akhir-akhir ini dikembangkan untuk meningkatkan performansi akurasi adalah SVM (*Support Vektor Machine*). SVM pertama kali dikenalkan oleh Cortes dan Vapnik (1995) sebagai hasil kolaborasi antara statistika dan *machine learning*. SVM merupakan salah satu metode yang baik untuk klasifikasi karena metode ini dapat dengan mudah mengatasi permasalahan klasifikasi berdimensi tinggi (Verplancke, et al., 2008). SVM memanfaatkan optimasi dengan *quadratic programming* yang apabila digunakan untuk data berdimensi tinggi dengan jumlah besar menjadi kurang efisien (Lee & Mangasarian, 2001). Salah satu teknik yang dikembangkan untuk mengatasi permasalahan ini adalah *smoothing technique* sehingga melahirkan metode baru yang disebut *Smooth Support Vector Machine* (SSVM). Berdasarkan penelitian Purnami dan Embong (2008), SSVM menghasilkan akurasi yang lebih baik bila dibandingkan dengan analisis diskriminan linier, *neural network*, *decision tree*, *genetic algorithm*, dan *supervised fuzzy clustering*.

Pengembangan metode SSVM terus dilakukan dengan memodifikasi *smoothing function*. Salah satunya adalah mengubah *smoothing function* standar ke dalam bentuk *polynomial smoothing function* yang memiliki beberapa formulasi, yaitu *quadratic polynomial function* dan *fourth polynomial function* yang digagas oleh Yuan dan Hang (2005), *piecewise polynomial function* oleh Luo, dkk. (2006), dan *spline function* oleh Yuan, dkk. (2007). Pada tahun 2009, Purnami, dkk. membandingkan keempat fungsi ini dan didapatkan hasil bahwa *piecewise polynomial function* memiliki pendekatan paling optimum terhadap *plus function* (Purnami, et al., 2009a) serta memiliki performansi akurasi yang lebih baik (Purnami, et al., 2009b). Pada tahun 2013, Wu dan Wang mengembangkan *piecewise polynomial function* yang memiliki formulasi berbeda dari *piecewise polynomial function* yang telah ditemukan oleh Luo, dkk. Pada penelitian ini, *piecewise polynomial function* yang digagas oleh Luo, dkk. akan disebut

sebagai *piecewise polynomial function 1* (PPSSVM1), sementara *piecewise polynomial function* yang ditemukan oleh Wu dan Wang disebut sebagai *piecewise polynomial function 2* (PPSSVM2). Pada tahun 2014, Wulandari membandingkan kedua fungsi *piecewise polynomial* tersebut dan diperoleh hasil bahwa secara teoritis *piecewise polynomial 2* memiliki performasi klasifikasi lebih baik bila dibandingkan dengan *piecewise polynomial 1* (Wulandari, 2014).

Berdasarkan keunggulan kedua fungsi *piecewise polynomial* tersebut maka penelitian ini akan mencoba mendapatkan model terbaik dari metode SSVM menggunakan kedua fungsi *piecewise polynomial function* untuk klasifikasi hasil *pap smear* pada Rumah Sakit “X” Surabaya.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan di atas maka permasalahan yang akan dibahas dalam penelitian ini adalah sebagai berikut.

1. Bagaimana karakteristik pasien berdasarkan faktor-faktor risiko penyebab kanker serviks?
2. Bagaimana evaluasi klasifikasi deteksi dini kanker serviks menggunakan *Smooth Support Vector Machine* (SSVM) berdasarkan fungsi *piecewise polynomial function 1* dan *piecewise polynomial function 2*?
3. Bagaimana model terbaik untuk klasifikasi deteksi dini kanker serviks berdasarkan performasi metode SSVM menggunakan fungsi *piecewise polynomial function 1* dan *piecewise polynomial function 2*?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendeskripsikan karakteristik pasien berdasarkan faktor-faktor risiko penyebab kanker serviks.
2. Mengevaluasi klasifikasi deteksi dini kanker serviks dengan menggunakan *Smooth Support Vector Machine* (SSVM)

berdasarkan fungsi *piecewise polynomial function 1* dan *piecewise polynomial function 2*.

3. Menemukan model terbaik untuk klasifikasi deteksi dini kanker serviks berdasarkan performasi metode SSVM menggunakan fungsi *piecewise polynomial function 1* dan *piecewise polynomial function 2*.

1.4 Manfaat Penelitian

Manfaat penelitian yang ingin dicapai adalah sebagai berikut.

1. Menambah wawasan pengetahuan perkembangan metode klasifikasi SSVM dan penerapannya dalam dunia kesehatan.
2. Menghasilkan teknik klasifikasi terbaik untuk klasifikasi deteksi dini kanker serviks menggunakan *Smooth Support Vektor Machine* (SSVM) antara *piecewise polynomial function 1* dan *piecewise polynomial function 2*.
3. Di bidang kesehatan penelitian ini membantu dokter untuk memberikan pertimbangan dalam mengklasifikasikan penyakit kanker serviks berdasarkan deteksi dini.

1.5 Batasan Masalah

Batasan dalam penelitian ini adalah sebagai berikut.

1. Fungsi kernel yang digunakan adalah fungsi kernel Gaussian.
2. Pembagian data *testing* dan *training* menggunakan *5-fold cross validation*.
3. *Smooth parameter* yang digunakan adalah $\alpha = 10$.
4. Luas AUC diestimasi menggunakan metode nonparametrik.



Halaman ini sengaja dikosongkan

BAB II

TINJAUAN PUSTAKA

2.1 *Support Vector Machine (SVM)*

Support Vector Machine (SVM) pertama kali diusulkan oleh Vapnik untuk klasifikasi dua kategori atau biner. Konsep SVM secara sederhana merupakan usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas, misalkan kelas +1 dan -1 pada *input space*. *Hyperplane* dalam ruang vektor berdimensi d adalah *affine subspace* berdimensi $d-1$ yang membagi ruang vektor tersebut ke dalam dua bagian yang masing-masing berkorespondensi pada kelas yang berbeda (Cortes & Vapnik, 1995).

Inti dari permasalahan klasifikasi biner adalah mengklasifikasikan suatu permasalahan ke dalam dua kelas. Misalkan diberikan satu set data terdiri atas m poin dalam ruang *real* berdimensi n . Data ini digambarkan sebagai matriks \mathbf{A} berukuran $m \times n$ di mana baris ke- i dari matriks \mathbf{A} , yaitu \mathbf{A}_i , berhubungan dengan poin data ke- i yang dinotasikan oleh sebuah vektor kolom $\mathbf{x} \in \mathbb{R}^n$. Kelas yang berlabel $y = \{+1\}$ atau $y = \{-1\}$ pada setiap poin data direpresentasikan pada diagonal matriks \mathbf{D} yang berukuran $m \times m$.

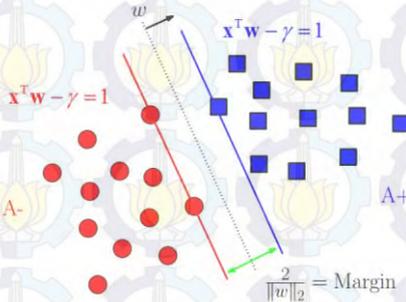
2.1.1 *Linearly Separable Case*

Pada klasifikasi biner, suatu data dikatakan *linearly separable* apabila kelas klasifikasi \mathbf{A}_+ dan \mathbf{A}_- dapat dengan tepat dipisahkan oleh *hyperplane*. Ilustrasi sederhana pada ruang *real* berdimensi dua dapat dilihat pada Gambar 2.1 di mana *hyperplane* yang terbentuk dapat memisahkan kedua kelas tanpa terjadi *missclassification*.

Pada permasalahan ini, kedua kelas dipisah dengan sepasang *bounding planes*, yaitu

$$\begin{aligned} \mathbf{x}^T \mathbf{w} - \gamma &= +1 \\ \mathbf{x}^T \mathbf{w} - \gamma &= -1 \end{aligned} \quad (2.1)$$

di mana vektor \mathbf{w} tegak lurus terhadap *bounding planes* (2.1) dan γ menentukan lokasi bidang pemisah terhadap titik asal. Vektor \mathbf{x} , \mathbf{w} berukuran $n \times 1$ yang menyusun ruang riil berdimensi R^n ($\mathbf{x}, \mathbf{w} \in R^n$) dan $\gamma \in R$.



Gambar 2.1 Ilustrasi dari *Linearly Separable Case* Pada SVM

Pada persamaan (2.1), *bounding plane* pertama membatasi kelas A_+ dan *bounding plane* kedua membatasi kelas A_- , sehingga.

$$\begin{aligned} \mathbf{x}^T \mathbf{w} - \gamma &\geq +1, \text{ untuk } \forall \mathbf{x} \in A_+ \\ \mathbf{x}^T \mathbf{w} - \gamma &\leq -1, \text{ untuk } \forall \mathbf{x} \in A_- \end{aligned} \quad (2.2)$$

Teori *statistical learning* yang dikemukakan oleh Vapnik (2000) pada Lee, dkk. (2012) menyatakan bahwa kemampuan SVM dalam memprediksi kelas klasifikasi dapat ditingkatkan dengan memaksimalkan $2/\|\mathbf{w}\|_2$ yang merupakan *margin* di antara dua *bounding planes*. Hal ini dapat dicapai dengan meminimumkan $\frac{1}{2}\|\mathbf{w}\|_2^2$. Formulasi dari permasalahan optimasi ini dapat ditulis menjadi.

$$\begin{aligned} \min_{(\mathbf{w}, b) \in R^{n+1}} & \frac{1}{2} \|\mathbf{w}\|_2^2 \\ \text{s.t. } & \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) \geq \mathbf{e} \end{aligned} \quad (2.3)$$

Sementara itu, bidang pemisah linier untuk kedua kelas adalah

$$\mathbf{x}^T \mathbf{w} = \gamma \quad (2.4)$$

yang terletak ditengah-tengah *bounding planes* (2.1) sebagaimana yang terlihat pada Gambar 2.1 (Lee, et al., 2012). Sehingga *classifier* dari SVM linier adalah (Santosa, 2007)

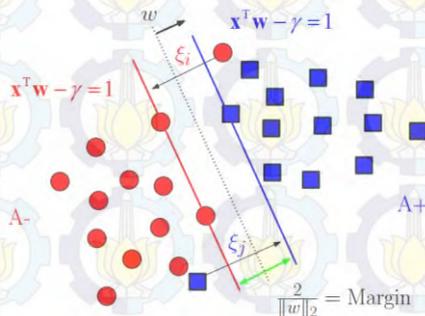
$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x})) \quad (2.5)$$

dengan

$$g(\mathbf{x}) = \mathbf{x}^T \mathbf{w} - \gamma \quad (2.6)$$

2.1.2 Linearly Nonseparable Case

Pada kenyataannya, suatu *hyperplane* yang dapat dengan sempurna memisahkan kedua kelas hampir tidak bisa ditemukan. Bila data bersifat *linearly nonseparable* maka *bounding planes* (2.2) tidak dapat digunakan karena akan membuat sampel data berada di kelas yang salah seperti yang ditampilkan pada Gambar 2.2.



Gambar 2.2 Ilustrasi dari Linearly Nonseparable Case Pada SVM

Untuk mengatasi kasus yang *linearly nonseparable* maka bidang pemisah kedua kelas ditentukan oleh sebuah “*soft margin*”, yaitu $\xi := (\xi_1, \dots, \xi_m)$ yang merupakan vektor *slack* nonnegatif berukuran $m \times 1$ sehingga

$$\begin{aligned} \mathbf{x}^T \mathbf{w} - \gamma + \xi_i &\geq +1 \text{ untuk } \mathbf{x}^T = \mathbf{A}_i \text{ dan } D_{ii} = +1 \\ \mathbf{x}^T \mathbf{w} - \gamma - \xi_i &\leq -1 \text{ untuk } \mathbf{x}^T = \mathbf{A}_i \text{ dan } D_{ii} = -1 \end{aligned} \quad (2.7)$$

$$(i = 1, 2, \dots, m)$$

Secara ringkas, konstrain (2.7) dapat ditulis sebagai sebuah persamaan matriks tunggal sebagai berikut

$$\mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \boldsymbol{\xi} \geq \mathbf{e} \quad (2.8)$$

di mana $\mathbf{e} \in \mathbb{R}^m$ adalah vektor berukuran $m \times 1$ dan bernilai 1.

Bentuk 1-norm dari vektor variabel *slack* $\boldsymbol{\xi}$, yaitu $\mathbf{e}^T \boldsymbol{\xi}$, disebut dengan *training error*. Bidang pemisah linier dengan *training error* yang paling kecil dapat ditemukan dengan menyelesaikan program linier sebagai berikut (Lee & Mangasarian, 2001).

$$\begin{aligned} \min_{(\mathbf{w}, \gamma, \boldsymbol{\xi}) \in \mathbb{R}^{n+1+m}} \mathbf{e}^T \boldsymbol{\xi} \\ \text{s.t. } \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \boldsymbol{\xi} \geq \mathbf{e}, \boldsymbol{\xi} \geq \mathbf{0} \end{aligned} \quad (2.9)$$

Namun, program linier (2.9) hanya berfokus untuk meminimumkan *training error* yang dapat menyebabkan *overfitting* pada data *training* di mana bidang pemisah dapat mengklasifikasikan data *training* dengan sangat baik namun kurang baik dalam mengklasifikasikan data *testing*. Permasalahan ini dapat diatasi dengan menggabungkan formulasi persamaan (2.3) dan persamaan (2.9) sehingga diperoleh formulasi SVM konvensional, yaitu

$$\begin{aligned} \min_{(\mathbf{w}, \gamma, \boldsymbol{\xi}) \in \mathbb{R}^{n+1+m}} \nu \mathbf{e}^T \boldsymbol{\xi} + \frac{1}{2} \|\mathbf{w}\|_2^2 \\ \text{s.t. } \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \boldsymbol{\xi} \geq \mathbf{e} \\ \boldsymbol{\xi} \geq \mathbf{0} \end{aligned} \quad (2.10)$$

di mana $\nu > 0$ adalah parameter positif yang menyeimbangkan bobot dari *training error* $\mathbf{e}^T \boldsymbol{\xi}$ dan *margin maximization term* $\frac{1}{2} \|\mathbf{w}\|_2^2$.

Fungsi objektif pada persamaan (2.10) merupakan permasalahan nonlinier. Secara komputasi, penyelesaian dari

permasalahan ini cukup sulit dan membutuhkan waktu yang lebih panjang. Oleh karena itu, fungsi primal dari SVM pada persamaan (2.10) ditransformasi ke dalam *dual space* untuk mempermudah penyelesaian optimasi sehingga lebih efisien dalam waktu komputasi (Santosa, 2007). Bentuk dual dari SVM yang disebut juga sebagai *Wolfe dual problem* memiliki formulasi sebagai berikut.

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{R}^m} \quad & \frac{1}{2} \mathbf{u}^T \mathbf{D} \mathbf{A} \mathbf{A}^T \mathbf{D} \mathbf{u} - \mathbf{e}^T \mathbf{u} \\ \text{s.t.} \quad & \mathbf{e}^T \mathbf{D} \mathbf{u} = 0, \quad 0 \leq \mathbf{u} \leq \mathbf{v} \end{aligned} \quad (2.11)$$

Variabel primal dari \mathbf{w} memiliki formulasi

$$\mathbf{w} = \mathbf{A}^T \mathbf{D} \mathbf{u} = \sum_{\{i|u_i > 0\}} u_i D_{ii} \mathbf{A}_i^T \quad (2.12)$$

dan γ diberikan oleh

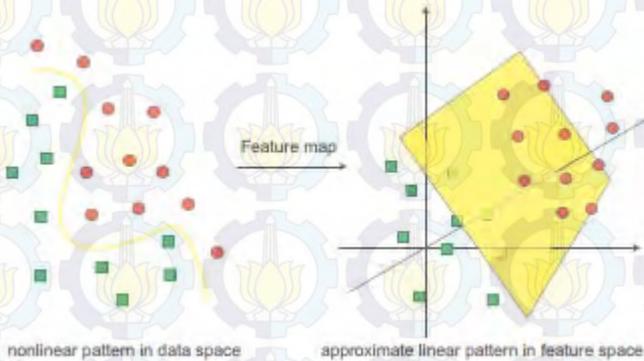
$$\gamma = \mathbf{A}_i \mathbf{w} - D_{ii} = \mathbf{A}_i \mathbf{A}^T \mathbf{D} \mathbf{u} - D_{ii}. \quad (2.13)$$

2.1.3 Nonlinear Case

Pada umumnya banyak data yang tidak dapat dipisahkan menggunakan bidang pemisah linier namun dapat dipisahkan secara linier bila dipetakan pada ruang dimensi yang lebih tinggi menggunakan sebuah pemetaan nonlinier. Pada kondisi yang demikian maka akan lebih baik menggunakan SVM nonlinier daripada menggunakan SVM linier. Pada SVM nonlinier, sebuah kernel akan digunakan untuk pemetaan nonlinier dan kemudian dicari *hyperplane* optimal menggunakan SVM linier. Hampir semua fungsi kernel nonlinier merupakan sebuah matriks yang berukuran besar sehingga membuat SVM nonlinier membutuhkan waktu komputasi yang lama saat memetakan data ke dalam dimensi yang lebih tinggi.

Misalkan data *training* dari *input space* \mathbb{R}^n dipetakan oleh sebuah *nonlinear map* \mathcal{Q} pada *feature space* yang memiliki dimensi lebih tinggi, yaitu F , maka data *training* \mathbf{A} pada F menjadi $\varphi(\mathbf{A}) \in \mathbb{R}^{m \times \ell}$ di mana ℓ adalah dimensi dari *feature space* F . Jika

inner product $\varphi(\mathbf{A})_i \varphi(\mathbf{A})_j^T$ diketahui untuk semua $i, j=1, 2, \dots, m$, maka algoritma SVM linier dapat dijalankan pada *feature space* F . *Hyperplane* yang terbentuk akan linier pada *feature space* F , namun akan membentuk permukaan yang nonlinier pada *input space* R^n seperti yang diilustrasikan pada Gambar 2.3 (Lee, et al., 2012).



Gambar 2.3 Ilustrasi dari *Nonlinear Case* pada SVM

Pada permasalahan ini, bentuk pemetaan nonlinier tidak perlu diketahui secara eksplisit. Pemetaan nonlinier ini dapat diperoleh dengan memanfaatkan sebuah fungsi kernel yang memenuhi kondisi *Mercer*, yaitu semidefinit positif. Oleh karena itu fungsi objektif SVM pada $\varphi(\mathbf{A})$ di dalam *feature space* F dengan mengganti kernel linier $\mathbf{A}\mathbf{A}^T$ pada persamaan (2.11) dengan kernel nonlinier $K(\mathbf{A}, \mathbf{A}^T) \in R^{m \times m}$. Hasil dari formulasi dual dari SVM nonlinier menjadi.

$$\begin{aligned} \min_{\mathbf{u} \in R^m} \quad & \frac{1}{2} \mathbf{u}^T \mathbf{D} \mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e}^T \mathbf{u} \\ \text{s.t.} \quad & \mathbf{e}^T \mathbf{D} \mathbf{u} = 0, \quad 0 \leq \mathbf{u} \leq \mathbf{ve} \end{aligned} \quad (2.14)$$

Bidang pemisah nonlinier yang dihasilkan adalah.

$$K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} = \gamma \quad (2.15)$$

di mana

$$\gamma = K(\mathbf{A}_i, \mathbf{A}^T) \mathbf{D} \mathbf{u} - D_{ii}, \quad i \in \mathbf{I} := \{j \mid 0 < u_j < v\} \quad (2.16)$$

Pemanfaatan fungsi kernel $K(\mathbf{A}, \mathbf{A}^T)$ yang bebas dari kondisi *Mercer* melahirkan *Generalized Support Vector Machine* (GSVM) dengan fungsi objektif sebagai berikut.

$$\min_{(\mathbf{u}, \gamma, \xi) \in \mathbb{R}^{m+1+m}} v \mathbf{e}^T \xi + f(\mathbf{u}) \quad (2.17)$$

Pada persamaan (2.17), $f(\mathbf{u})$ adalah fungsi konveks pada \mathbb{R}^m yang menekan parameter \mathbf{u} untuk memaksimalkan jarak antara *bounding planes*. Sementara parameter $v > 0$ memboboti *error* klasifikasi sejalan dengan proses *suppression* dari parameter \mathbf{u} . Sebuah solusi dari program matematika untuk \mathbf{u} dan γ menghasilkan bidang pemisah nonlinier, yaitu.

$$K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} = \gamma \quad (2.18)$$

Beberapa macam fungsi kernel yang dapat digunakan untuk pemetaan nonlinier, yaitu (Mangasarian, 1998)

a. Kernel *Polynomial*

$$K(\mathbf{A}, \mathbf{A}^T) = (\mathbf{A} \mathbf{A}^T + \mu \mathbf{r} \mathbf{r}^T)^d \quad (2.19)$$

b. Kernel *Neural Network*

$$K(\mathbf{A}, \mathbf{A}^T) = (\mathbf{A} \mathbf{A}^T + \mu \mathbf{r} \mathbf{r}^T) \quad (2.20)$$

c. Kernel *Gaussian*

$$K(\mathbf{A}, \mathbf{A})_{ij} = \exp\left(-\mu \|\mathbf{A}_i^T - \mathbf{A}_j^T\|^2\right) \quad (2.21)$$

di mana dalam hal ini μ , r , dan d merupakan parameter kernel dengan $i, j = 1, 2, \dots, m$ dan $\ell = m$.

Hsu, dkk. (2008) menganjurkan fungsi kernel Gaussian sebagai pilihan pertama dan utama dalam pemilihan *kernel trick*. Hal ini dikarenakan kemampuan fungsi kernel Gaussian yang dapat memetakan sampel nonlinier ke dalam sebuah ruang dimensi

yang lebih tinggi sehingga bisa menangani kasus ketika hubungan antara kelas label dan atribut tidak linier. Alasan berikutnya adalah jumlah *hyperparameter* yang mempengaruhi kerumitan *model selection*. Kernel ini memiliki kesulitan numerik lebih kecil karena bila dibandingkan kernel *polynomial*, kernel Gaussian memiliki *hyperparameter* yang lebih sedikit, yaitu μ .

2.2 Smooth Support Vector Machine (SSVM)

Smooth Support Vector Machine (SSVM) merupakan formulasi baru dari SVM yang menggunakan *smoothing technique*. SSVM memiliki sifat matematika yang penting yaitu kecembungan yang kuat dan terdiferensialkan tak hingga. Berdasarkan sifat-sifat ini, solusi unik dari SSVM akan konvergen terhadap solusi unik problem optimasi SVM bila *smooth parameter* α pada SSVM mendekati tak hingga (Lee & Mangasarian, 2001).

2.2.1 SSVM Linier

Pada pendekatan *smoothing* yang digagas oleh Lee dan Mangasarian (2001), kuadrat *2-norm* dari vektor variabel *slack* ξ diminimalkan dengan bobot $\nu/2$ menggantikan *1-norm* dari vektor variabel *slack* ξ pada persamaan (2.10). Sebagai tambahan, jarak antara *bounding planes* (2.1) diukur pada ruang dimensi $(n+1)$ dari $(\mathbf{w}, \gamma) \in R^{n+1}$, yaitu $2/\|(\mathbf{w}, \gamma)\|_2$ sehingga hasil modifikasi dari fungsi objektif SVM linier standar (2.10) adalah sebagai berikut

$$\begin{aligned} \min_{(\mathbf{w}, \gamma, \xi) \in R^{n+1+m}} & \frac{\nu}{2} \xi^T \xi + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2) \\ \text{s.t} & \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) + \xi \geq \mathbf{e}, \xi \geq \mathbf{0} \end{aligned} \quad (2.22)$$

Kendala pada persamaan (2.22) dapat ditulis menjadi

$$\xi = (\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ \quad (2.23)$$

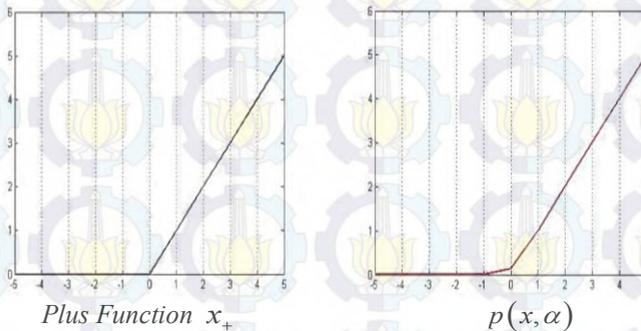
di mana dalam hal ini $x_+ = \max\{0, x\}$. Bila persamaan (2.23) disubstitusikan ke dalam persamaan (2.22) maka akan diperoleh fungsi objektif bebas kendala, yaitu.

$$\min_{(\mathbf{w}, \gamma) \in \mathbb{R}^{n+1}} \frac{\nu}{2} \left\| (\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ \right\|_2^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2) \quad (2.24)$$

Bila dibandingkan dengan fungsi objektif (2.22), fungsi objektif (2.24) mereduksi dimensi dari yang semula $n+1+m$ menjadi $n+1$. Namun, fungsi objektif ini tidak memiliki turunan kedua sehingga metode optimasi konvensional tidak bisa digunakan. Oleh karena itu, Lee dan Mangasarian (2001) mengusulkan *smoothing technique*, yaitu sebuah teknik di mana *plus function* x_+ pada fungsi objektif (2.24) didekati dengan sebuah fungsi *smooth-p* yang merupakan integral dari fungsi *sigmoid neural networks* $(1 + \exp(-\alpha x))^{-1}$ atau dapat ditulis sebagai berikut

$$p(x, \alpha) = x + \frac{1}{\alpha} \ln(1 + \exp(-\alpha x)) \quad (2.25)$$

dengan $\alpha > 0$ adalah *smooth parameter* yang mengontrol kedekatan kurva $p(x, \alpha)$ terhadap kurva *plus function* x_+ . Pada Gambar 2.4 terlihat bahwa *smoothing function* $p(x, \alpha)$ dapat mendekati *plus function* x_+ .



Gambar 2.4 Smooth Approximation untuk Plus Function

Bila *plus function* x_+ pada persamaan (2.24) diganti dengan $p(x, \alpha)$ maka diperoleh formulasi SSVM sebagai berikut.

$$\min_{(\mathbf{w}, \gamma) \in R^{n+1}} \Psi_{\alpha}(\mathbf{w}, \gamma) := \min_{(\mathbf{w}, \gamma) \in R^{n+1}} \frac{\nu}{2} \|p(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)), \alpha\|_2^2 + \frac{1}{2} (\mathbf{w}^T \mathbf{w} + \gamma^2) \quad (2.26)$$

Fungsi objektif SSVM (2.26) memiliki turunan kedua sehingga problem optimasi dapat dilakukan dengan menerapkan algoritma konvergen kuadrat Newton dengan tahapan (*stepsize*) Armijo yang membuat algoritma tersebut konvergen secara global (Lee & Mangasarian, 2001).

Algoritma Newton-Armijo dimulai dengan menetapkan *starting value* $(\mathbf{w}^0, \gamma^0) \in R^{n+1}$. Proses akan berhenti apabila gradien dari fungsi objektif SSVM, yaitu $\nabla \Psi_{\alpha}(\mathbf{w}^i, \gamma^i) = 0$. Selain itu, $(\mathbf{w}^{i+1}, \gamma^{i+1})$ akan dihitung dengan langkah-langkah sebagai berikut.

- (i) **Newton Direction** : Menentukan *direction* $d^i \in R^{n+1}$ dengan menetapkan linierisasi sama dengan nol dari $\nabla \Psi_{\alpha}(\mathbf{w}, \gamma)$ di sekitar (\mathbf{w}^i, γ^i) yang memberikan $n+1$ persamaan linier dengan $n+1$ variabel.

$$\nabla^2 \Psi_{\alpha}(\mathbf{w}^i, \gamma^i) d^i = -\nabla \Psi_{\alpha}(\mathbf{w}^i, \gamma^i)' \quad (2.27)$$

- (ii) **Armijo Stepsize [1]** : Memilih sebuah *stepsize* $\lambda_i \in R$ sedemikian hingga

$$(\mathbf{w}^{i+1}, \gamma^{i+1}) = (\mathbf{w}^i, \gamma^i) + \lambda_i d^i \quad (2.28)$$

di mana $\lambda_i = \max \left\{ 1, \frac{1}{2}, \frac{1}{4}, \dots \right\}$ sehingga

$$\Psi_{\alpha}(\mathbf{w}^i, \gamma^i) - \Psi_{\alpha}((\mathbf{w}^i, \gamma^i) + \lambda_i d^i) \geq -\delta \lambda_i \nabla \Psi_{\alpha}(\mathbf{w}^i, \gamma^i) d^i \quad (2.29)$$

dengan $\delta \in \left(0, \frac{1}{2} \right)$.

Saat $\nabla\Psi_\alpha(\mathbf{w}^i, \gamma^i) = 0$, iterasi pada algoritma Newton-Armijo berhenti dan diperoleh nilai \mathbf{w} dan γ yang konvergen. Pada proses optimasi ini, matriks Hessian diberikan oleh (Lee & Mangasarian, 2001)

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} \nabla^2 \Psi_\alpha(\mathbf{w}, \gamma) &= \begin{bmatrix} \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \gamma)}{\partial^2 \mathbf{w}^2} & \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \gamma)}{\partial \mathbf{w} \partial \gamma} \\ \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \gamma)}{\partial \gamma \partial \mathbf{w}} & \frac{\partial^2 \Psi_\alpha(\mathbf{w}, \gamma)}{\partial^2 \gamma^2} \end{bmatrix} \\ &= I + \nu \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{bmatrix} \end{aligned} \quad (2.30)$$

di mana

$$\mathbf{H}_{11} = \mathbf{A}^T \text{diag}(S_\infty(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)))\mathbf{A}$$

$$\mathbf{H}_{12} = -\mathbf{A}^T \text{diag}(S_\infty(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)))\mathbf{e}$$

$$\mathbf{H}_{21} = -\mathbf{e}^T \text{diag}(S_\infty(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)))\mathbf{A}$$

$$\mathbf{H}_{22} = \mathbf{e}^T \text{diag}(S_\infty(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)))\mathbf{e}$$

dan

$$S_\infty(\mathbf{x}) = \lim_{\alpha \rightarrow \infty} \frac{1}{1 + \varepsilon^{-\alpha \mathbf{x}}} = \frac{1 + \text{sign}(\mathbf{x})}{2}.$$

Sementara itu, gradien dari matriks Hessian memiliki formulasi sebagai berikut (Lee & Mangasarian, 2001).

$$\lim_{\alpha \rightarrow \infty} \nabla \Psi_\alpha(\mathbf{w}, \gamma) = \begin{bmatrix} \mathbf{w} - \nu \mathbf{A}^T \mathbf{D}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ \\ \gamma + \nu \mathbf{e}^T \mathbf{D}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ \end{bmatrix} \quad (2.31)$$

Nilai \mathbf{w} dan γ optimum kemudian membentuk *classifier* dari SSVM linier yang memiliki formulasi (Lee & Mangasarian, 2001).

$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x})) \quad (2.32)$$

dengan

$$g(\mathbf{x}) = \mathbf{x}^T \mathbf{w} - \gamma \quad (2.33)$$

2.2.2 SSVM Nonlinier

Formulasi SSVM nonlinier dikembangkan dari formulasi GSVM (2.17) yang menggunakan fungsi kernel untuk menentukan bidang pemisah nonlinier. Pada modifikasi formulasi GSVM ini juga digunakan kuadrat dari 2-norm variabel *slack* ξ dengan pembobot $\nu/2$ sebagai *training error* sehingga diperoleh persamaan kuadratik sebagai berikut

$$\begin{aligned} \min_{(\mathbf{u}, \gamma, \xi) \in \mathbb{R}^{m+1+m}} & \frac{\nu}{2} \xi^T \xi + \frac{1}{2} (\mathbf{u}^T \mathbf{u} + \gamma^2) \\ \text{s.t. } & \mathbf{D} \left(\mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e} \gamma \right) + \xi \geq \mathbf{e}, \quad \xi \geq \mathbf{0} \end{aligned} \quad (2.34)$$

Kendala pada persamaan (2.34) dapat ditulis menjadi.

$$\xi = \left(\mathbf{e} - \mathbf{D} \left(\mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e} \gamma \right) \right)_+ \quad (2.35)$$

Bila Persamaan (2.35) dimasukkan ke dalam Persamaan (2.34) maka akan diperoleh problem minimisasi bebas kendala sebagai berikut.

$$\min_{(\mathbf{u}, \gamma) \in \mathbb{R}^{m+1}} \frac{\nu}{2} \left\| \left(\mathbf{e} - \mathbf{D} \left(\mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e} \gamma \right) \right)_+ \right\|_2^2 + \frac{1}{2} (\mathbf{u}^T \mathbf{u} + \gamma^2) \quad (2.36)$$

Pada Persamaan (2.36), *plus function* x_+ diganti dengan fungsi *smooth* $p(x, \alpha)$ sehingga diperoleh fungsi objektif SSVM nonlinier sebagai berikut.

$$\begin{aligned} \min_{(\mathbf{u}, \gamma) \in \mathbb{R}^{m+1}} & \Psi_\alpha(\mathbf{u}, \gamma) = \\ & \min_{(\mathbf{u}, \gamma) \in \mathbb{R}^{m+1}} \frac{\nu}{2} \left\| p \left(\mathbf{e} - \mathbf{D} \left(\mathbf{K}(\mathbf{A}, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \mathbf{e} \gamma \right), \alpha \right) \right\|_2^2 + \frac{1}{2} (\mathbf{u}^T \mathbf{u} + \gamma^2) \end{aligned} \quad (2.37)$$

Bidang pemisah nonlinier ditentukan oleh solusi optimum dari \mathbf{u} dan γ dari (2.37) yang dapat diselesaikan dengan algoritma Newton Armijo sehingga *decision function* $f(\mathbf{x})$ yang terbentuk untuk SSVM nonlinier adalah sebagai berikut (Lee & Mangasarian, 2001)

$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x})) \quad (2.38)$$

dengan

$$g(\mathbf{x}) = K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma \quad (2.39)$$

2.2.3 Contoh Implementasi Algoritma Newton Armidjo Pada SSVM

Berikut ini akan diberikan contoh implementasi algoritma Newton Armidjo, penerapan *classifier* pada kasus linier dan nonlinier.

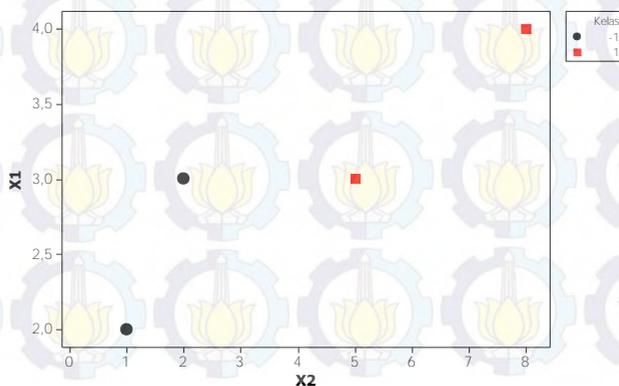
A. Implementasi Algoritma Newton Armidjo pada SSVM Linier

Misalkan diberikan sebuah data seperti pada Tabel 2.1 berikut ini.

Tabel 2.1 Data Ilustrasi Newton-Armidjo

Y	1	1	-1	-1
X1	3	4	2	3
X2	5	8	1	2

Bila data klasifikasi pada Tabel 2.1 diplotkan dalam sebuah bidang *cartesius* maka hasilnya nampak seperti pada Gambar 2.5 berikut ini.



Gambar 2.5 Scatter Plot Data Ilustrasi Newton Armidjo

Berdasarkan Gambar 2.5, terlihat bahwa kedua kelas dapat dipisahkan oleh bidang pemisah linier (*linear case*).

Misalkan dari hasil seleksi parameter diperoleh $\nu=2$ dan $\gamma=0$. Tahapan-tahapan dari algoritma Newton-Armidjo adalah sebagai berikut.

1) *Inisialisasi Data*

$$\mathbf{A} = \begin{bmatrix} 3 & 5 \\ 4 & 8 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

2) *Menghitung Gradien Matriks Hessian*

Initial point untuk menghitung gradien matriks Hessian menggunakan formulasi (2.31) adalah

$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ dan } \gamma=0$$

Menggunakan *initial point* tersebut, dihitung $\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)$. Hasilnya adalah

$$\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \left\{ \begin{bmatrix} 3 & 5 \\ 4 & 8 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

sehingga

$$\begin{aligned} \mathbf{w} - \nu \mathbf{A}^T \mathbf{D}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 2 \begin{bmatrix} 3 & 4 & 2 & 3 \\ 5 & 8 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -4 \\ -20 \end{bmatrix} \end{aligned}$$

$$\gamma + \mathbf{v}^T \mathbf{D}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))_+ = 0 + 2 \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 0$$

$$\lim_{\alpha \rightarrow \infty} \nabla \Psi_{\alpha}(\mathbf{w}, \gamma) = \begin{bmatrix} -4 \\ -20 \\ 0 \end{bmatrix}$$

3) Mengecek Gradien > 0

$$\|\nabla \Psi_{\alpha}(\mathbf{w}, \gamma)\|_2^2 = \begin{bmatrix} -4 & -20 & 0 \end{bmatrix} \begin{bmatrix} -4 \\ -20 \\ 0 \end{bmatrix} = 416$$

Ternyata, gradien > 0 ($416 > 0$) sehingga proses dilanjutkan pada tahapan Newton *Direction*.

4) Menghitung Gradien dari Matriks Hessian

Pada tahapan Newton *Direction*, matriks Hessian dihitung terlebih dahulu menggunakan formulasi (2.30).

$$S_{\infty}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma)) = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{diag}(S_{\infty}(\mathbf{e} - \mathbf{D}(\mathbf{A}\mathbf{w} - \mathbf{e}\gamma))) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{H}_{11} = \begin{bmatrix} 3 & 4 & 2 & 3 \\ 5 & 8 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 4 & 8 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 38 & 55 \\ 55 & 94 \end{bmatrix}$$

$$\mathbf{H}_{12} = - \begin{bmatrix} 3 & 4 & 2 & 3 \\ 5 & 8 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -12 \\ -16 \end{bmatrix}$$

$$\mathbf{H}_{21} = - \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 4 & 8 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} -12 & -16 \end{bmatrix}$$

$$\mathbf{H}_{22} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 4$$

Matriks Hessian yang terbentuk adalah sebagai berikut.

$$\lim_{\alpha \rightarrow \infty} \nabla^2 \Psi_{\alpha}(\mathbf{w}, \gamma) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + 2 \begin{bmatrix} 38 & 55 & -12 \\ 55 & 94 & -16 \\ -12 & -16 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} 77 & 110 & -24 \\ 110 & 189 & -32 \\ -24 & -32 & 9 \end{bmatrix}$$

5) Menentukan Newton Direction (d^i)

Misalkan $\mathbf{H} = \nabla^2 \Psi_{\alpha}(\mathbf{w}^i, \gamma^i)$ dan $\mathbf{G} = \nabla \Psi_{\alpha}(\mathbf{w}^i, \gamma^i)$ maka

$$d^i = \mathbf{H}^{-1}(-\mathbf{G}) = \begin{bmatrix} 77 & 110 & -24 \\ 110 & 189 & -32 \\ -24 & -32 & 9 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 20 \\ 0 \end{bmatrix} = \begin{bmatrix} -0,52 \\ 0,44 \\ 0,16 \end{bmatrix}$$

6) Menentukan Stepsize dan Memperbaharui Initial Point

Tahapan awal penentuan *stepsize* adalah dengan menghitung persamaan (2.29) di mana $\Psi_\alpha(\mathbf{w}, \gamma)$ dapat dilihat pada persamaan (2.26).

$$\Psi_\alpha(\mathbf{w}^i, \gamma^i) = \Psi_\alpha \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) = \frac{2}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [1 \ 1 \ 1] + 0 = 4.$$

$$\Psi_\alpha((\mathbf{w}^i, \gamma^i) + \lambda_i d^i) = \Psi_\alpha \left(\begin{bmatrix} -0,52 \\ 0,44 \\ 0,16 \end{bmatrix} \right) = 0,16.$$

$$-\delta \lambda \nabla \Psi_\alpha(\mathbf{w}, \gamma)^T d = -\frac{1}{2}(1) [-4 \ -20 \ 0] \begin{bmatrix} -0,52 \\ 0,44 \\ 0,16 \end{bmatrix} = 3,36.$$

Sehingga

$$\Psi_\alpha(\mathbf{w}^i, \gamma^i) - \Psi_\alpha((\mathbf{w}^i, \gamma^i) + \lambda_i d^i) \geq -\delta \lambda_i \nabla \Psi_\alpha(\mathbf{w}^i, \gamma^i)^T d^i$$

$$\Leftrightarrow (4 - 0,16) > 3,36$$

$$\Leftrightarrow 3,84 > 3,36.$$

Perbarui (\mathbf{w}, γ) dengan menghitung persamaan (2.28).

Hasilnya adalah sebagai berikut.

$$(\mathbf{w}^i, \gamma^i) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + (1) \begin{bmatrix} -0,52 \\ 0,44 \\ 0,16 \end{bmatrix} = \begin{bmatrix} -0,52 \\ 0,44 \\ 0,16 \end{bmatrix}$$

Proses akan terus diulangi sampai gradien < 0 .

B. Penerapan SSVM Pada Kasus Linier

Misalkan hasil optimasi menggunakan algoritma Newton Armidjo pada poin A mendapatkan hasil \mathbf{w} dan γ optimum sebagai berikut

$$\mathbf{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \text{ dan } \gamma = 1$$

maka *classifier* yang terbentuk adalah $f(\mathbf{x}) = \text{sign}(g(\mathbf{x}))$ dengan

$$g(\mathbf{x}) = \mathbf{x}^T \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 1.$$

Bila suatu amatan baru yang memiliki nilai $x_1 = 2$ dan $x_2 = 1$ maka

$$g(\mathbf{x}) = \begin{bmatrix} 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} - 1 = 6.$$

Karena $f(\mathbf{x}) = \text{sign}(6) = 1$ maka amatan baru tersebut masuk dalam kelas 1.

C. Penerapan SSVM Pada Kasus Nonlinier

Misalkan data pada merupakan data *training* yang digunakan untuk pembentukan model SSVM nonlinier.

Tabel 2.2 Contoh Data *Training*

Y	X ₁	X ₂
-1	3	2
-1	2	1
1	1	3

Maka susunan matriks pada algoritma SSVM nonlinier adalah.

$$\mathbf{D} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Misalkan dari hasil penghitungan menggunakan algoritma SSVM nonlinier diperoleh.

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \text{ dan } \gamma = 1$$

Terdapat satu pengamatan baru yang ingin dicari kelas klasifikasinya, yaitu.

$$\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Maka proses pengklasifikasian data baru tersebut adalah sebagai berikut.

1) Mendefinisikan fungsi kernel Gaussian $K(\mathbf{x}^T, \mathbf{A}^T)$

Pada contoh ini \mathbf{x}^T berukuran 1×2 dan \mathbf{A} berukuran 3×2 sehingga kernel gaussian $K(\mathbf{x}^T, \mathbf{A}^T)$ berukuran 1×3 .

$$K(\mathbf{x}^T, \mathbf{A}^T) = [K_{11} \quad K_{12} \quad K_{13}]$$

$$K_{ij} = \exp\{-\mu[\mathbf{d}_{ij} * \mathbf{d}_{ij}^T]\}$$

dengan

$$\mathbf{d}_{ij} = \mathbf{x}_i^T - \mathbf{A}_j,$$

$$i = 1 \text{ dan } j = 1, 2.$$

Misalkan $\mu = 0,1$.

- Untuk $i=1$ dan $j=1$

$$\mathbf{d}_{11} = \mathbf{x}^T - \mathbf{A}_1 = [0 \quad 1] - [3 \quad 2] = [-3 \quad -1]$$

$$\mathbf{d}_{11} * \mathbf{d}_{11}^T = [-3 \quad -1] \begin{bmatrix} -3 \\ -1 \end{bmatrix} = 10$$

$$K_{11} = \exp\{-\mu[\mathbf{d}_{11} * \mathbf{d}_{11}^T]\} = \exp\{-0,1(10)\} = \exp(-1) = 0,3679$$

- Untuk $i=1$ dan $j=2$

$$\mathbf{d}_{12} = \mathbf{x}^T - \mathbf{A}_2 = [0 \ 1] - [2 \ 1] = [-2 \ 0]$$

$$\mathbf{d}_{12} * \mathbf{d}_{12}^T = [-2 \ 0] \begin{bmatrix} -2 \\ 0 \end{bmatrix} = 4$$

$$K_{12} = \exp\{-\mu[\mathbf{d}_{12} * \mathbf{d}_{12}^T]\} = \exp\{-0,1(4)\} = \exp(-0,4) = 0,6703$$

- Untuk $i=1$ dan $j=3$

$$\mathbf{d}_{13} = \mathbf{x}^T - \mathbf{A}_3 = [0 \ 1] - [1 \ 3] = [-1 \ -2]$$

$$\mathbf{d}_{13} * \mathbf{d}_{13}^T = [-1 \ -2] \begin{bmatrix} -1 \\ -2 \end{bmatrix} = 5$$

$$K_{13} = \exp\{-\mu[\mathbf{d}_{13} * \mathbf{d}_{13}^T]\} = \exp\{-0,1(5)\} = \exp(-0,5) = 0,6065$$

Dari penghitungan di atas, diperoleh

$$K(\mathbf{x}^T, \mathbf{A}^T) = [0,3679 \ 0,6703 \ 0,6065]$$

- 2) Melakukan klasifikasi menggunakan *decision function* yang terbentuk

Classifier yang terbentuk adalah sebagai berikut

$$g(\mathbf{x}) = K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma = 0,1111 - 1 = -0,8889$$

sehingga $f(\mathbf{x}) = \text{sign}(g(\mathbf{x})) = \text{sign}(-0,8889) = -1$.

Hasil dari *decision function* mengklasifikasikan amatan baru \mathbf{x} ke dalam kelas -1.

2.3 Polynomial Smoothing Function

Salah satu hasil modifikasi *smooth function* yang dikembangkan oleh beberapa peneliti adalah *polynomial smoothing function*. Berikut ini beberapa jenis dari *polynomial smoothing function*.

1. *Quadratic Polynomial Function*

Fungsi ini memiliki formulasi sebagai berikut (Yuan & Huang, 2005).

$$f_1(x, \alpha) = \begin{cases} x, & x > \frac{1}{\alpha} \\ \frac{\alpha}{4} + \frac{1}{2}x + \frac{1}{4\alpha}, & -\frac{1}{\alpha} \leq x \leq \frac{1}{\alpha} \\ 0, & x < -\frac{1}{\alpha} \end{cases} \quad (2.40)$$

2. Fourth Polynomial Function

Formulasi dari fungsi ini adalah sebagai berikut (Yuan, et al., 2005).

$$f_2(x, \alpha) = \begin{cases} x, & x > \frac{1}{\alpha} \\ -\frac{\alpha^3}{16} \left(x + \frac{1}{\alpha}\right)^3 \left(x - \frac{3}{\alpha}\right), & -\frac{1}{\alpha} \leq x \leq \frac{1}{\alpha} \\ 0, & x < -\frac{1}{\alpha} \end{cases} \quad (2.41)$$

3. Spline Function

Persamaan *spline function* adalah sebagai berikut (Yuan, et al., 2007).

$$f_3(x, \alpha) = \begin{cases} 0, & x < -\frac{1}{\alpha} \\ \frac{\alpha^2}{6}x^3 + \frac{\alpha}{2}x^2 + \frac{1}{2}x + \frac{1}{6\alpha}, & -\frac{1}{\alpha} \leq x < 0 \\ -\frac{\alpha^2}{6}x^3 + \frac{\alpha}{2}x^2 + \frac{1}{2}x + \frac{1}{6\alpha}, & 0 \leq x \leq \frac{1}{\alpha} \\ x, & \frac{1}{\alpha} < x \end{cases} \quad (2.42)$$

4. Piecewise Polynomial Function

Terdapat dua versi mengenai rumusan *piecewise polynomial function*, yaitu.

a. *Piecewise Polynomial Function 1*

Persamaan *piecewise polynomial function 1* diberikan oleh (Luo, et al., 2006).

$$f_4(x, \alpha) = \begin{cases} x, & x \geq \frac{1}{\alpha} \\ \frac{1}{32}(\alpha x + 1)^3, & -\frac{1}{\alpha} < x < \frac{1}{\alpha} \\ \left(\alpha^2 x^3 - 3\alpha x^2 + x + \frac{5}{\alpha} \right), & \\ 0, & x \leq -\frac{1}{\alpha} \end{cases} \quad (2.43)$$

b. *Piecewise Polynomial Function 2*

Persamaan *piecewise polynomial function 2* adalah sebagai berikut (Wu & Wang, 2013).

$$f_5(x, \alpha) = \begin{cases} 0, & x < -\frac{1}{3\alpha} \\ \frac{3}{2}\alpha^2 \left(x + \frac{1}{3\alpha} \right)^3, & -\frac{1}{3\alpha} \leq x < 0 \\ x + \frac{3}{2}\alpha^2 \left(\frac{1}{3\alpha} - x \right), & 0 \leq x \leq \frac{1}{3\alpha} \\ x, & x > \frac{1}{3\alpha} \end{cases} \quad (2.44)$$

Pada penelitian ini, *smooth parameter* yang digunakan adalah $\alpha = 10$ dengan acuan penelitian Wu dan Wang klasifikasi (2013) mengenai *Piecewise Smooth Support Vector Machine* untuk klasifikasi.

2.4 Seleksi Paramater

Permasalahan dalam pemilihan sebuah parameter yang baik untuk performasi generalisasi yang lebih baik dalam *learning task* disebut *model selection*. Salah satu metode standar yang dapat

digunakan untuk *model selection* adalah menggunakan *exhaustive grid search* sederhana atas ruang parameter. Namun, metode ini tidak dapat secara efektif melakukan *model selection* secara otomatis karena biaya komputasi yang tinggi. Oleh karena itu, banyak perbaikan metode yang diusulkan untuk mengurangi jumlah percobaan dalam kombinasi parameter.

Pada tahun 2007, Huang, dkk. mengembangkan suatu metode untuk *model selection* pada SVM yang memungkinkan pengguna menemukan kombinasi parameter yang baik secara efisien dan otomatis, disebut *Uniform Design* (UD). Metode ini telah dicoba pada beberapa *learning tasks* dan set data yang berbeda-beda, serta pada beberapa algoritma SVM. Hasil perhitungan numerik dan perbandingan dengan metode-metode lain menunjukkan bahwa *Uniform Design* dapat secara efektif menemukan sebuah kombinasi parameter yang baik dengan hanya membutuhkan sedikit percobaan. Skema pemilihan model ini *robust* dan efisien serta dapat dilakukan secara otomatis (Huang, et al., 2007).

Pada metode SVM, seleksi parameter dilakukan terhadap parameter fungsi kernel dan parameter penyeimbang SVM, yaitu ν . Fungsi kernel yang paling sering digunakan dalam pemetaan nonlinier adalah kernel gaussian yang memiliki parameter μ . Kedua parameter ini menjadi kunci performansi SVM. Ketika nilai μ terlalu besar maka risiko terjadinya *overfitting* semakin besar. Bila nilai μ terlalu kecil maka risiko terjadinya *underfitting* semakin besar. Sementara itu, parameter ν dari SVM berfungsi untuk meminimumkan *training error* dan mengurangi kekompleksan model (Huang, et al., 2007).

Huang, dkk. (2007) menyatakan bahwa nilai dari sebuah parameter kernel gaussian yang baik berada dalam rentang

$$\frac{10^{-3}}{\rho} < \mu < \frac{1,9}{\rho} \quad (2.45)$$

di mana $\rho = \min_{A_i \neq A_j} \|A_i^T - A_j\|$. Ketika $\mu < 10^{-3} / \rho$ maka akan terjadi kondisi *underfitting*, sementara bila $\mu > 1,9 / \rho$ maka akan terjadi kondisi *overfitting*. Di sisi lain, sebuah parameter penyeimbang v yang baik berada dalam rentang

$$10^{-2} \leq v \leq 10^4 \quad (2.46)$$

2.5 Evaluasi Performasi Klasifikasi

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performasi teknik klasifikasi secara keseluruhan. Semakin tinggi akurasi klasifikasi berarti performasi teknik klasifikasi juga semakin baik.

Untuk permasalahan klasifikasi biner, akurasi klasifikasi dapat dilihat melalui tabel klasifikasi. Dari tabel klasifikasi dapat diuji tingkat sensitivitas dan spesifisitas. Sensitivitas adalah tingkat akurasi kelas positif sementara spesifisitas adalah tingkat akurasi kelas negatif. Berikut ini tabel klasifikasi seperti yang ditunjukkan pada Tabel 2.3.

Tabel 2.3 Tabel Klasifikasi

Kondisi yang Sebenarnya	Hasil Prediksi		
	Positif	Negatif	Total Baris
Positif	TP	FN	TP+FN (Total kelas positif yang sebenarnya)
Negatif	FP	TN	FP+TN (Total kelas negatif yang sebenarnya)
Total Kolom	TP+FP (Total prediksi kelas positif)	FN+TN (Total prediksi kelas negatif)	N=TP+TN+FP+FN (Total diagnosa)

Ada beberapa istilah yang umum digunakan bersamaan dengan deskripsi sensitivitas, spesifisitas, dan akurasi, yaitu *True*

Positive (TP), *True Negative* (TN), *False Negative* (FN), dan *False Positive* (FP). Keterangan dari masing-masing istilah tersebut adalah sebagai berikut (Zhu, et al., 2010).

- TP : *True Positive* (hasil prediksi dan kondisi sebenarnya sama-sama positif)
- TN : *True Negative* (hasil prediksi dan kondisi yang sebenarnya sama-sama negatif)
- FP : *False Positive* (hasil prediksi positif namun sebenarnya negatif)
- FN : *False Negative* (hasil prediksi negatif namun sebenarnya positif)

Penghitungan akurasi, sensitivitas, dan spesifisitas adalah sebagai berikut.

a. Akurasi

$$\text{Akurasi} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (2.47)$$

b. Sensitivitas

$$\text{Sensitivitas} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.48)$$

c. Spesifisitas

$$\text{Spesifisitas} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.49)$$

2.5.1 *Cross Validation*

Cross validation adalah metode yang sering digunakan untuk evaluasi model. Metode *holdout* adalah jenis yang paling sederhana dari *cross validation*. Pada metode ini, kumpulan data dipisahkan menjadi dua set, yaitu data *training* dan data *testing*. Data *training* berguna untuk mengestimasi model, sementara data *testing* digunakan untuk menguji model yang dibentuk oleh data *training*. Metode ini biasanya lebih disukai daripada metode residual dan tidak membutuhkan waktu yang lama dalam perhitungannya. Namun, proses evaluasinya dapat memiliki variansi yang tinggi. Evaluasi sangat bergantung pada data mana yang masuk ke dalam data *training* dan data *testing*. Dengan

demikian, hasil evaluasi mungkin akan jauh berbeda tergantung pada pembagian data ke dalam set *training* maupun *testing*.

K-fold cross validation adalah salah satu cara untuk meningkatkan performansi metode *holdout*. Kumpulan data dibagi menjadi k subset dan metode *holdout* diulang sebanyak k kali. Setiap pengulangan, salah satu dari k subset digunakan untuk data *testing* dan sisanya, $k-1$ subset, digunakan untuk membentuk model (data *training*). Kemudian rata-rata *error* dari semua k percobaan dihitung. Metode ini tidak lagi memperlmasalahakan bagaimana data dibagi menjadi dua set karena semua subset akan menjadi data *training* maupun *testing* sehingga variansi dari hasil estimasi berkurang. Namun, metode ini membutuhkan waktu komputasi yang lebih lama karena algoritma *cross validation* harus diulang dari awal sebanyak k kali (Schneider, 1997).

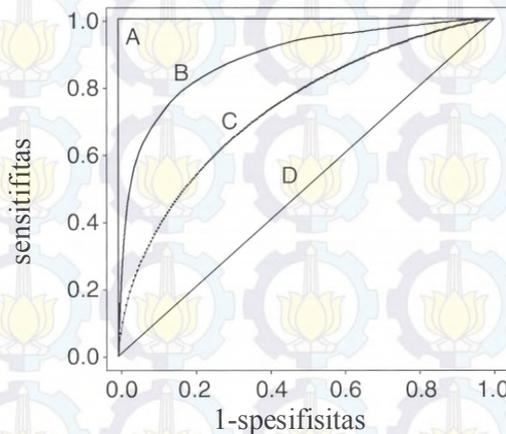
2.5.2 Kurva *Receiving Operating Characteristic* (ROC)

Sensitivitas dan spesifisitas merupakan salah satu ukuran dasar yang digunakan untuk mengevaluasi performansi klasifikasi dalam dunia medis. Sebuah model klasifikasi biner hanya memiliki sepasang nilai sensitivitas dan spesifisitas. Bila suatu permasalahan klasifikasi dievaluasi menggunakan beberapa metode maka pemilihan model terbaik akan sulit dilakukan karena terdapat beberapa pasang nilai sensitivitas dan spesifisitas. Untuk mengatasi permasalahan ini maka digunakan kurva ROC untuk mengukur performansi setiap model klasifikasi. Kurva ROC (*Receiver Operating Characteristics*) adalah representasi grafis dari hubungan timbal balik antara sensitivitas dan spesifisitas (Erke & Pattynama, 1998) di mana kurva ini sering digunakan untuk mengevaluasi pengklasifikasian karena mempunyai kemampuan secara menyeluruh dan cukup baik (Chou, et al., 2010).

Pada kurva ROC, *true positif rate* (sensitivitas) diplot dalam fungsi *false positif rate* (1-spesifisitas) untuk poin *cut off* yang berbeda-beda. Setiap titik pada plot ROC merupakan pasangan sensitivitas dan spesifisitas yang sesuai dengan batasan keputusan tertentu. Sebuah tes dengan diskriminasi sempurna memiliki plot ROC yang melewati sudut kiri atas (sensitivitas 100%, spesifisitas

100%). Oleh karena itu, semakin dekat plot ROC ke sudut kiri atas, semakin tinggi akurasi keseluruhan tes (Zweig & Campbell, 1993).

Metode yang umum digunakan untuk menghitung performansi klasifikasi adalah dengan menghitung luas daerah di bawah kurva ROC, yaitu *Area Under The ROC Curve* (AUC). Nilai AUC berada di antara 0 dan 1. Bila nilai AUC semakin mendekati 1 maka model klasifikasi yang diperoleh lebih akurat. Kurva ROC yang baik berada di daerah sebelah atas dari garis diagonal (0,0) dan (1,1) sehingga tidak ada nilai AUC yang lebih kecil dari 0,5.



Gambar 2.6 Empat Kurva ROC dengan Nilai AUC yang Berbeda

Gambar 2.6 menunjukkan empat kurva ROC, yaitu A, B, C, dan D yang memiliki nilai AUC berbeda-beda. Sebuah tes yang sempurna memiliki nilai AUC sebesar 1, digambarkan oleh kurva A. Sementara itu, kurva D yang membentuk garis diagonal dari koordinat (0,0) sampai dengan (1,1) memiliki nilai AUC sebesar 0,5. Kurva ROC dari sebuah tes klasifikasi yang memiliki kemampuan untuk membedakan antara kelompok positif dan negatif berada di antara dua kurva ekstrim A dan D, yaitu kurva B dan C. Dalam hal ini, kurva B memiliki luas area di bawah kurva ROC lebih besar dari kurva C sehingga model klasifikasi yang

diwakili oleh kurva B memiliki performasi klasifikasi lebih baik dari model klasifikasi yang diwakili oleh kurva C.

Terdapat dua metode dalam mengestimasi luasan AUC, yaitu metode parametrik dan nonparametrik. Asumsi utama dalam mengestimasi AUC menggunakan metode parametrik adalah hasil dari tes diagnosis biner pada kelompok positif (Y_1) dan kelompok negatif (Y_0) masing-masing berdistribusi normal dengan rata-rata populasi yang berbeda. Sementara pada metode nonparametrik, asumsi distribusi tentang pengukuran tes diagnostik tidak dibutuhkan (Colak, et al., 2012). Oleh karena banyak peneliti yang lebih suka menggunakan metode nonparametrik daripada metode parametrik.

Misalkan Y merupakan sebuah variabel random yang mewakili hasil tes diagnosis biner, yaitu 0 dan 1, maka diagnosis berdasarkan sembarang *cutoff* c adalah positif bila $Y \geq c$ dan negatif jika $Y < c$. Nilai *cutoff* c dihitung menggunakan rumusan sebagai berikut

$$\begin{aligned}c_1 &= kk - 1 \\c_2 &= \frac{kk + kb}{2} \\c_3 &= kb + 1\end{aligned} \quad (2.50)$$

dengan kk adalah kelas terkecil (0) dan kb adalah kelas terbesar (1) pada hasil tes diagnosis.

Penghitungan sensitivitas (TP) dan 1-spesifisitas (FP) pada sembarang *cutoff* c diberikan oleh

$$\begin{aligned}TP(c) &= \frac{s_1(c)}{n_1} \\FP(c) &= \frac{s_0(c)}{n_0}\end{aligned} \quad (2.51)$$

di mana

$s_1(c)$: jumlah subjek $Y \geq c$ pada kelas 1

$s_0(c)$: jumlah subjek $Y \geq c$ pada kelas 0

n_1 : jumlah kejadian kelas 1

n_0 : jumlah kejadian kelas 0.

Estimasi AUC menggunakan metode nonparametrik dihitung menggunakan *trapezoidal rule* dengan rumusan sebagai berikut (Colak, et al., 2012)

$$AUC = \frac{1}{n_1 n_0} \sum_{i=1}^{n_1} \sum_{j=1}^{n_0} \varphi(Y_{i1}, Y_{j0}) \quad (2.52)$$

di mana

$$\varphi(Y_{i1}, Y_{j0}) = \begin{cases} 1 & \text{jika } Y_{i1} > Y_{j0} \\ 1/2 & \text{jika } Y_{i1} = Y_{j0} \\ 0 & \text{jika } Y_{i1} < Y_{j0} \end{cases} \quad (2.53)$$

serta Y_{i1} dan Y_{j0} masing-masing merupakan hasil tes diagnostik untuk kelas positif dan negatif.

Kategori pengklasifikasian model berdasarkan nilai AUC diberikan oleh Tabel 2.4.

Tabel 2.4 Kategori Pengklasifikasian Model Berdasarkan Nilai AUC

Nilai AUC	Model Diklasifikasikan Sebagai
0,90-1,00	<i>Excellent</i>
0,80-0,90	<i>Good</i>
0,70-0,80	<i>Fair</i>
0,60-0,70	<i>Poor</i>
0,50-0,60	<i>Fail</i>

Sumber : Colak, et al. (2007)

Misalkan kelas asli dan hasil klasifikasi biner diberikan oleh Tabel 2.5.

Tabel 2.5 Hasil Klasifikasi Biner

Kelas Asli	1	1	1	0	0	0
Prediksi	1	0	1	0	0	0

Kelas terbesar pada data tersebut adalah 1, sementara kelas terkecil adalah 0 sehingga *cutoff* (c) yang terbentuk adalah

$$c_1 = 0 - 1 = -1$$

$$c_2 = \frac{0+1}{2} = 0,5$$

$$c_3 = 1+1 = 2.$$

Prediksi kelas berdasarkan ketiga nilai *cutoff* ini diberikan oleh Tabel 2.6.

Tabel 2.6 Prediksi Kelas Berdasarkan Nilai *Cutoff*

Kelas Asli	Hasil Prediksi	$c = -1$	$c = 0,5$	$c = 2$
1	1	1	1	0
1	0	1	0	0
1	1	1	1	0
0	0	1	0	0
0	0	1	0	0
0	0	1	0	0

Hasil prediksi kelas pada Tabel 2.6 kemudian digunakan untuk menentukan koordinat titik kurva ROC menggunakan persamaan (2.51). Hasilnya dapat dilihat pada Tabel 2.7.

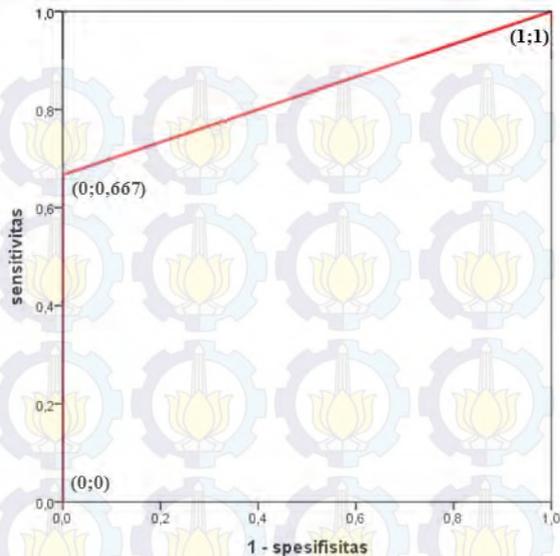
Tabel 2.7 Koordinat Titik Kurva ROC

c	Sensitivitas	1-Spesifisitas
-1	1	1
0,5	0,667	0
2	0	0

Kurva ROC yang terbentuk dapat dilihat pada Gambar 2.7. Sementara itu luas AUC dapat dihitung menggunakan *trapezoidal rule* sebagai berikut

$$\frac{0,667+1}{2} \times 1 = 0,8335.$$

Luas AUC yang dihasilkan adalah sebesar 0,8335 sehingga model yang terbentuk masuk dalam kategori *good model*.



Gambar 2.7 Kurva ROC yang Terbentuk

2.6 Kanker Serviks

Leher rahim adalah bagian dari sistem reproduksi wanita. Letaknya berada di panggul, merupakan bagian sempit dari rahim yang terletak di bagian paling bawah.

Kanker serviks tumbuh diawali dari sel tubuh. Sel-sel tubuh bersatu membentuk jaringan. Jaringan-jaringan yang berkumpul menjadi satu akhirnya membentuk leher rahim dan organ tubuh yang lain. Umumnya, sel leher rahim yang normal akan tumbuh dan beregenerasi sesuai dengan kebutuhan tubuh. Ketika sel-sel leher rahim mulai tua atau rusak, sel-sel tersebut akan mati dan kemudian digantikan oleh sel-sel baru. Terkadang, proses regenerasi ini tidak berjalan lancar. Sebagai contoh, sel-sel baru tumbuh ketika tubuh tidak membutuhkannya atau bisa jadi sel-sel yang sudah tua dan rusak tidak mati dan tidak mengalami regenerasi. Kegagalan proses regenerasi inilah yang memicu tumbuhnya tumor.

Pertumbuhan tumor pada leher rahim dapat berupa tumor jinak (bukan kanker) dan tumor ganas (kanker) dengan karakteristik sebagai berikut.

1. Tumor jinak
 - Tidak berbahaya
 - Tidak menjalar ke jaringan di sekitar tumor tersebut.
2. Tumor ganas
 - Dapat menyebabkan kematian
 - Dapat menjalar ke jaringan dan organ sekitar
 - Dapat menyebar ke bagian tubuh lain

Kanker serviks mulai menyebar dari permukaan leher rahim. Seiring berjalannya waktu, kanker ini dapat menginfeksi lebih jauh ke dalam leher rahim dan jaringan terdekat. Sel-sel kanker dapat menyebar dengan melepaskan diri dari tumor serviks. Sel-sel tumor tersebut akan melalui pembuluh limpa ke kelenjar getah bening dan juga dapat menyebar melalui pembuluh darah ke paru-paru, hati, maupun tulang. Setelah menyebar, sel kanker akan melekat pada jaringan lain dan tumbuh membentuk tumor baru yang dapat merusak jaringan tersebut (National Cancer Institute, 2012).

2.6.1 Faktor Risiko

Faktor risiko kanker serviks adalah faktor yang meningkatkan risiko menderita kanker serviks. Salah satu faktor risiko yang paling sering ditemukan pada kejadian kanker serviks adalah infeksi virus HPV. Sebagian besar manusia dewasa telah terinfeksi HPV, namun kebanyakan infeksi tersebut telah dinetralkan dan disembuhkan secara otomatis oleh sistem pertahanan tubuh. Infeksi HPV yang tidak disembuhkan dapat menyebabkan kanker serviks pada beberapa wanita.

Selain infeksi HPV, terdapat faktor risiko lain yang meningkatkan peluang seseorang terjangkit kanker serviks, yaitu.

1. Usia
Kanker serviks sering terjadi pada wanita usia 40 tahun ke atas dan sangat jarang terjadi pada wanita dengan usia kurang dari 15 tahun. Kanker ini juga dapat menyerang

manula. Kebanyakan manula berpikir bahwa *Pap Smear Test* tidak perlu lagi dilakukan setelah mengalami menopause (Nurwijaya, et al., 2010)

2. Siklus menstruasi

Menstruasi normal umumnya berlangsung selama 4-7 hari. Biasanya siklus menstruasi terjadi setiap 28 hari dan masih dianggap normal bila muncul setiap 21-35 hari. Siklus menstruasi tidak dianggap normal bila terdapat indikasi sebagai berikut.

- Menstruasi muncul kurang dari 21 hari atau lebih dari 35 hari
- Telat selama tiga siklus atau lebih berturut-turut
- Keluarnya darah menstruasi lebih berat atau lebih ringan dari biasanya
- Menstruasi yang terjadi lebih dari tujuh hari
- Menstruasi yang dibarengi dengan rasa sakit, kram, mual, dan muntah
- Pendarahan yang terjadi selama periode, setelah *menopause*, dan ketika berhubungan intim.

Banyak faktor yang menyebabkan siklus menstruasi tidak normal, salah satu di antaranya adalah gangguan pada leher rahim (Cleveland Clinic, 2010). Oleh karena itu siklus menstruasi yang tidak normal sering dikaitkan dengan kanker serviks.

3. Usia pertama kali menstruasi

Usia pertama kali menstruasi pada penderita kanker serviks rata-rata berada pada kisaran $14,5 \pm 0,71$ tahun (Yakasai, et al., 2013).

4. Usia pertama kali melahirkan

Risiko kanker serviks 77% lebih tinggi pada wanita yang pertama kali melahirkan di bawah 17 tahun dibandingkan dengan wanita yang berumur 25 tahun atau lebih. Risiko berkurang seiring dengan semakin bertambahnya usia saat pertama kali melahirkan (Cancer Research UK, 2014).

5. Paritas

Paritas adalah jumlah anak yang telah dilahirkan oleh seorang wanita baik hidup maupun mati. Penelitian yang dilakukan oleh Fatmawati pada tahun 2013 menemukan bahwa terdapat kaitan antara paritas lebih dari dua dengan kejadian kanker serviks. Risiko meningkat 1,32 kali setiap paritas bertambah satu baik pada kelompok usia risiko tinggi (≥ 45 tahun) maupun risiko rendah (< 45 tahun) (Fatmawati, 2012).

6. Penggunaan alat kontrasepsi

Penggunaan alat kontrasepsi oral yang dipakai dalam jangka waktu panjang dapat meningkatkan risiko relatif terjangkit kanker serviks 1,53 kali. WHO melaporkan risiko relatif pada pemakaian kontrasepsi oral sebesar 1,19 kali dan meningkat dengan lamanya pemakaian (Sjamsuddin, 2001). Kontrasepsi oral dapat meningkatkan risiko 1,5-2,5 kali bila diminum dalam jangka panjang, yaitu lebih dari 4 tahun (Dalimartha, 2004). Penelitian mengenai kaitan antara penggunaan alat kontrasepsi dengan kanker serviks juga dilakukan oleh yang mendapatkan hasil bahwa proporsi kejadian kanker serviks lebih banyak pada kelompok kasus yang memakai alat kontrasepsi hormonal dibandingkan dengan yang tidak memakai alat kontrasepsi hormonal. Wanita yang memakai alat kontrasepsi hormonal berisiko menderita kanker serviks 1,244 kali lebih besar dibanding ibu yang tidak memakai alat kontrasepsi hormonal (Tira, 2008).

7. Riwayat keguguran

Menggugurkan kandungan di dalam dunia kedokteran dikenal dengan istilah *abortus* yang berarti pengeluaran hasil konsepsi (pertemuan sel telur dan sperma) sebelum janin dapat hidup di luar kandungan. Wanita yang pernah mengalami riwayat keguguran ≥ 1 kali berisiko 3,37 kali lebih besar untuk menderita kanker serviks dibandingkan wanita yang tidak pernah mengalami keguguran (Abbas,

2003). Penelitian yang digagas oleh Tira pada tahun 2008 menemukan bahwa proporsi kejadian kanker serviks lebih banyak pada ibu yang riwayat keguguran ≥ 1 kali sebesar 84,5% dibandingkan dengan yang tidak mempunyai riwayat keguguran sebesar 15,5%. Serta wanita yang mempunyai riwayat keguguran ≥ 1 kali berisiko menderita kanker serviks 7,713 kali lebih besar dibandingkan wanita yang tidak pernah keguguran (Tira, 2008).

2.6.2 *Pap Smear Test*

Pada tahun 1928, dr.Papanicolaou menemukan bahwa sel-sel dalam leher rahim mengalami perubahan bentuk sebelum menjadi kanker. *Pap smear* merupakan pemeriksaan sitologi yang digunakan untuk mengecek perubahan pada leher rahim di bagian atas vagina. Tes ini merupakan alat *screening* untuk menemukan gejala dini mengenai perkembangan sel kanker di masa mendatang.

Pap smear termasuk prosedur yang sederhana. Sel dikumpulkan dari leher rahim dan ditempatkan (doleskan) ke *slide*. *Slide* kemudian dikirim ke laboratorium untuk menguji sampel sel yang didapatkan. Jika perubahan abnormal ditemukan pada *screening*, tes lebih lanjut akan dilakukan untuk melihat apakah pengobatan dibutuhkan (Australian Department of Health, 2013)

Hasil *pap smear test* sesuai klasifikasi *Papanicolou* terdiri dari lima kelas, yaitu (Nurwijaya, et al., 2010).

1. Normal, yaitu tidak ditemukan sel abnormal.
2. Tampak sel abnormal tetapi tidak ada bukti keganasan.
3. Tampak sel-sel *atypik* yang meragukan untuk keganasan.
4. Tampak sel-sel yang mengindikasikan keganasan.
5. Tampak adanya sel-sel ganas.

2.7 Penelitian Sebelumnya

Beberapa penelitian mengenai klasifikasi hasil deteksi dini kanker serviks menggunakan *Pap Smear Test* pernah digagas sebelumnya oleh Yastuti menggunakan metode *Bagging Multivariate Adaptive Regression Splines* (MARS) untuk klasifikasi hasil *pap smear* kanker serviks di Rumah Sakit “X”

Surabaya dan mendapatkan ketepatan klasifikasi sebesar 79,67% (Yastuti, 2011). Tahun berikutnya, Intansari melakukan penelitian dengan kasus yang sama menggunakan metode *Bagging Logistic Regression* dan diperoleh hasil ketepatan klasifikasi sebesar 66,7% dari hasil regresi logistik biner yang kemudian meningkat menjadi 70,74% setelah menggunakan pendekatan *bagging* regresi logistik (Intansari, 2012).

Sementara itu metode PPSSVM telah digunakan pada beberapa penelitian berbasis klasifikasi dan terbukti memiliki tingkat performansi yang tinggi. Pada tahun 2009, Purnami, dkk. menggunakan keempat fungsi *polynomial*, yaitu *piecewise polynomial function 1*, *spline function*, *fourth polynomial function* dan *quadratic polynomial function* untuk klasifikasi kanker payudara dan terbukti bahwa SSVM dengan fungsi *piecewise polynomial 1* memiliki akurasi paling baik, yaitu sebesar 98,59% untuk data *training* dan 98,25% untuk data *testing* (Purnami, et al., 2009b). Sementara itu, Wulandari (2014) mengaplikasikan kedua fungsi *piecewise polynomial* untuk klasifikasi desa tertinggal di Provinsi Kalimantan Timur tahun 2011 dan diperoleh hasil bahwa akurasi klasifikasi untuk *piecewise polynomial 1* adalah sebesar 89,45% dan 89,59% untuk *piecewise polynomial 2*.

BAB III METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari Rumah Sakit “X” Surabaya bagian Riset dan Pengembangan tahun 2010. Data mengenai hasil *pap smear test* dan faktor-faktor nonmedis yang diduga mempengaruhi penyakit kanker serviks berjumlah 3586 kasus dengan rincian jumlah hasil tes yang normal sebanyak 2414 kasus dan abnormal sebanyak 1172 kasus.

3.2 Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini terdiri atas variabel respon dan variabel prediktor. Variabel respon yang terdiri atas dua kelas merupakan hasil *pap smear test* berdasarkan gambaran sitologi. Sementara variabel prediktor merupakan faktor risiko kanker serviks yang terdiri atas tujuh faktor. Sebelumnya, pemilihan dan pembentukan variabel-variabel ini didapatkan dari hasil diskusi dengan pihak Rumah Sakit “X” Surabaya.

Tabel 3.1 Gambaran Sitologi Hasil *Pap Smear Test*

Kelompok	Jenis	
Normal	Normal <i>smear</i>	
Abnormal	Radang non spesifik	TV
	AS (<i>Atrophic Smear</i>)	HV
	ASC-US	Candida
	ASC-H	Coccen
	Early as	LSIL
	HSIL	Mild RNS
	Herpes Simplek Virus	Metaplasia

Gambaran hasil pemeriksaan sitologi dari *pap smear test* di Rumah Sakit “X” Surabaya ditampilkan pada Tabel 3.1. Hasil ini kemudian dibagi menjadi dua kelompok, yaitu kelompok normal dan abnormal. Kelompok normal mengindikasikan bahwa tidak

terjadi perubahan pada sampel sel leher rahim, sementara kelompok abnormal mengindikasikan terdapat perubahan pada sampel sel leher rahim. Kedua kelompok ini kemudian dijadikan variabel respon (Y) dengan ketentuan sebagai berikut.

Y = 0 : normal

Y = 1 : abnormal

Sementara itu variabel prediktor (X) yang digunakan adalah sebagai berikut.

Tabel 3.2 Variabel Prediktor

Variabel		Deskripsi	Skala
X ₁	Usia	Usia dari pasien	Rasio
X ₂	Usia pertama kali menstruasi	-	Rasio
X ₃	Usia pertama kali melahirkan	-	Rasio
X ₄	Siklus menstruasi	1 : Teratur 2 : Tidak Teratur	Nominal
X ₅	Paritas	1 : ≤ 2 anak 2 : > 2 anak	Nominal
X ₆	Penggunaan alat kontrasepsi	1 : Tidak 2 : Ya	Nominal
X ₇	Riwayat keguguran	1 : Pernah 2 : Tidak pernah	Nominal

Berikut adalah struktur data penelitian yang disajikan pada Tabel 3.3.

Tabel 3.3 Struktur Data Penelitian

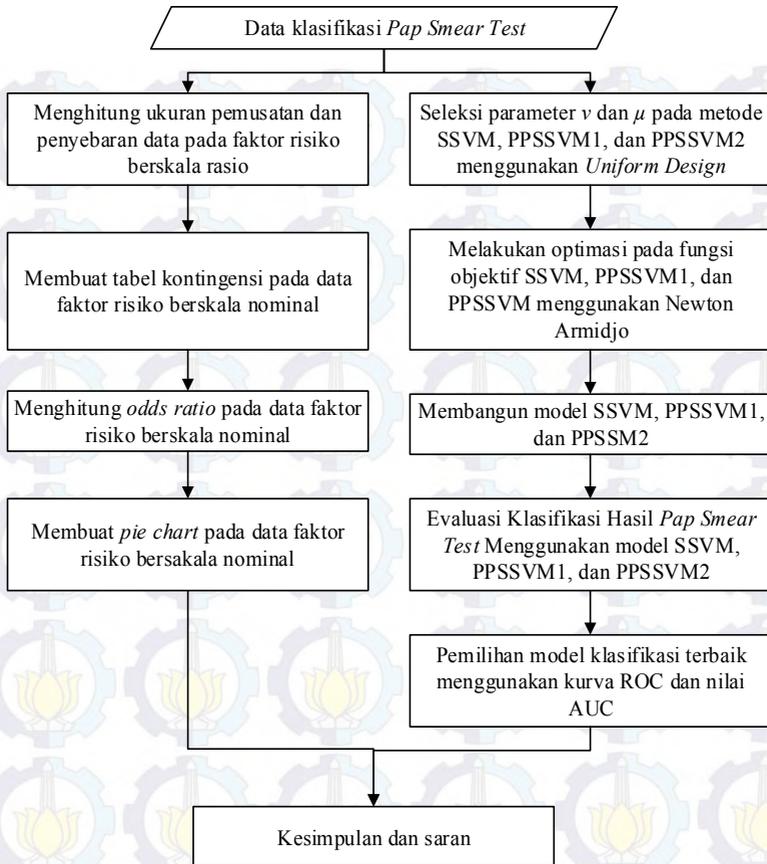
Hasil <i>Pap Smear Test</i>	Y	X ₁	X ₂	...	X ₇
1	y ₁	x ₁₁	x ₂₁	...	x ₇₁
2	y ₂	x ₁₂	x ₂₁	...	x ₇₁
⋮	⋮	⋮	⋮	⋮	⋮
N	y _n	x ₁₃	x _{1n}	...	x _{7n}

3.3 Metode Analisis

Terdapat enam tahapan untuk melakukan analisis data dalam penelitian ini, yaitu.

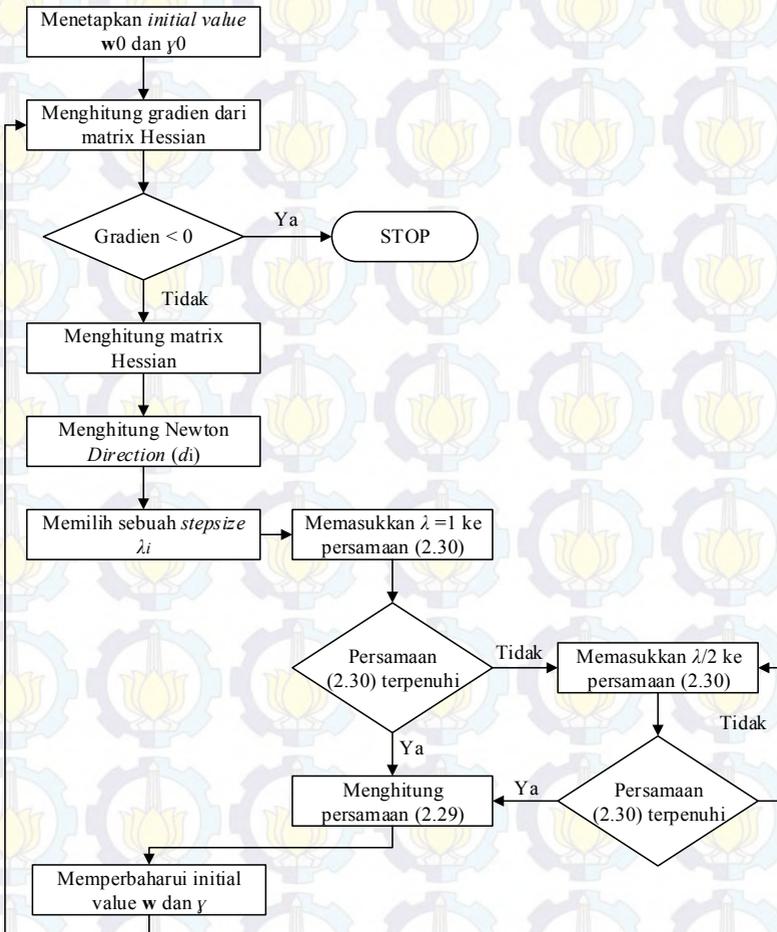
1. Menjelaskan karakteristik pasien *pap smear test* berdasarkan faktor risiko yang mempengaruhi kanker serviks.
 - Menghitung ukuran pemusatan dan penyebaran data pada data faktor risiko dengan skala rasio.
 - Membuat tabel kontingensi dan *pie chart*, serta menghitung *odds ratio* pada data faktor risiko berskala nominal.
2. Melakukan seleksi parameter pada fungsi kernel Gaussian dan parameter penyeimbang SVM menggunakan *Uniform Design* pada metode SSVM, PPSSVM1, dan PPSSVM2.
3. Menyelesaikan permasalahan optimasi SSVM, PPSSVM1, dan PPSSVM2 menggunakan Newton Armidjo.
4. Membangun model SSVM, PPSSVM1, dan PPSSVM2 untuk klasifikasi deteksi dini kanker serviks.
5. Melakukan evaluasi performasi model SSVM dan PPSSVM yang telah didapatkan untuk klasifikasi deteksi dini kanker serviks. Langkah-langkahnya adalah sebagai berikut.
 - Membagi data menjadi dua bagian, yaitu data *training* dan data *testing* dengan menggunakan *5-fold cross validation*.
 - Menghitung akurasi, sensitivitas, dan spesifisitas.
 - Membuat kurva ROC.
 - Menghitung nilai AUC.
6. Menentukan model terbaik antara SSVM, PPSSVM1, dan PPSSVM2 untuk klasifikasi deteksi dini kanker serviks.

Langkah-langkah yang telah dipaparkan sebelumnya dapat diringkas dalam bentuk kerangka penelitian seperti yang disajikan dalam Gambar 3.1.



Gambar 3.1 Tahapan Penelitian

Sementara itu, algoritma Newton Armidjo pada tahapan optimasi SSVM, PPSSVM1, dan PPSSVM2 diberikan oleh Gambar 3.2.



Gambar 3.2 Algoritma Newton Armidjo



Halaman ini sengaja dikosongkan

BAB IV

ANALISIS DAN PEMBAHASAN

4.1 Karakteristik Pasien Berdasarkan Faktor-Faktor Risiko Penyebab Kanker Serviks

Pada pembahasan awal, gambaran karakteristik pasien yang melakukan *pap smear test* akan diberikan berdasarkan ketujuh faktor risiko dari kanker serviks. Faktor risiko yang memiliki skala rasio akan dihitung ukuran pemusatan dan penyebaran datanya, sementara untuk faktor risiko yang memiliki skala nominal akan digambarkan secara visual disertai tabulasi silang dan *odds ratio*. Data yang digunakan pada tahapan ini merupakan data asli hasil *pap smear test* sebelum dilakukan metode *balancing* data.

4.1.1 Karakteristik Pasien Berdasarkan Faktor Usia, Usia Pertama Kali Menstruasi, dan Usia Pertama Kali Melahirkan

Pap smear test merupakan salah satu metode skrining untuk mendeteksi perubahan pada sel-sel lahir rahim yang dapat mengindikasikan penyakit kanker serviks. Ketika hasil *pap smear* menunjukkan adanya perubahan abnormal pada sampel sel leher rahim maka dokter akan melakukan pemeriksaan lanjutan seperti pemeriksaan kolposkopi (pemeriksaan leher rahim dengan mengoleskan asam asetat di leher rahim) dan tes HPV (*Human Papilloma Virus*) untuk mengetahui ada tidaknya infeksi HPV yang merupakan faktor risiko utama dari kanker serviks (Anonim, 2014).

Hasil dari *pap smear test* sendiri dapat diprediksi melalui faktor-faktor nonmedis yang berkaitan dengan kanker serviks. Pada penelitian ini digunakan tujuh faktor risiko dengan tiga faktor diantaranya memiliki skala rasio, yaitu faktor usia, usia pertama kali melahirkan, dan usia pertama kali menstruasi. Pada Tabel 4.1 akan ditunjukkan karakteristik pasien *pap smear* pada Rumah Sakit “X” Surabaya tahun 2010 secara deskriptif berdasarkan ketiga faktor risiko tersebut.

Tabel 4.1 Statistika Deskriptif Data Pasien *Pap Smear Test*

	Usia	Usia Pertama Kali Menstruasi	Usia Pertama Kali Melahirkan
N	3586	3586	3586
Minimum	14	8	12
Maksimum	80	19	49
Rata-Rata	42,378	12,935	25,812
Deviasi Standar	9,179	1,447	4,008
Modus	39	12	25
N Modus	201	1105	418

Berdasarkan Tabel 4.1, pasien yang melakukan *pap smear test* di Rumah Sakit “X” Surabaya pada tahun 2010 rata-rata berumur sekitar 42 tahun dengan usia 14 tahun dan 80 tahun sebagai usia termuda dan tertua pasien *pap smear*. Sementara itu, karakteristik pasien yang paling banyak melakukan *pap smear test* adalah wanita berusia 39 tahun dengan kejadian sebanyak 201, pertama kali menstruasi pada umur 12 tahun sebanyak 1105 kejadian, dan melahirkan pada umur 25 tahun untuk pertama kalinya sejumlah 418 kejadian. Dari ketiga faktor risiko ini, faktor usia memiliki deviasi standar terbesar yang menandakan bahwa pasien yang melakukan *pap smear test* memiliki variasi usia yang lebih besar bila dibandingkan dengan kedua karakteristik yang lain.

Selain mengetahui karakteristik pasien *pap smear test* secara umum, penting pula diketahui karakteristik pasien dengan hasil *pap smear test* normal maupun abnormal seperti yang diberikan oleh Tabel 4.2. Berdasarkan Tabel 4.2, pasien yang menunjukkan hasil *abnormal* pada *pap smear test* rata-rata berumur 40 tahun dengan kejadian terbanyak pada pasien dengan umur 39 tahun. Hal ini sesuai dengan penelitian yang dilakukan oleh *Surveillance Epidemiology and End Results of US National Cancer Institute* (2011) yang menunjukkan bahwa rata-rata penyakit kanker serviks terbanyak didiagnosis di antara umur 35 dan 44 tahun. Meskipun demikian, karakteristik pada kelompok hasil *pap smear* abnormal ini tidak jauh berbeda bila dibandingkan dengan kelompok hasil

pap smear normal yang memiliki rata-rata usia pasien sebesar 43 tahun dan kejadian terbanyak pada usia 40 tahun. Bila dilihat dari deviasi standar pada kedua kelompok ini, kelompok hasil *pap smear* normal memiliki variasi usia pasien yang lebih tinggi dibandingkan dengan kelompok hasil *pap smear* abnormal.

Tabel 4.2 Statistika Deskriptif Pasien Pada Kelompok Hasil *Pap Smear*

	Usia		Usia Pertama Kali Menstruasi		Usia Pertama Kali Melahirkan	
	(N)	(A)	(N)	(A)	(N)	(A)
N	2414	1172	2414	1172	2414	1172
Minimum	20	14	8	9	12	12
Maksimum	80	71	19	18	49	45
Rata-Rata	43,348	40,380	12,975	12,853	25,638	26,172
Deviasi Standar	10,000	6,776	1,482	1,370	3,985	4,032
Modus	40	39	12	12	25	27
N Modus	122	84	750	355	296	140

*(N) = hasil normal; (A) = hasil abnormal

Pada penelitian ini, pasien *pap smear test* dengan hasil normal maupun abnormal pada Rumah Sakit “X” Surabaya tahun 2010 rata-rata mengalami menstruasi untuk yang pertama kalinya pada usia sekitar 12 tahun dengan jumlah kejadian terbanyak yang juga berada pada usia 12 tahun. Sementara untuk karakteristik berdasarkan usia pertama kali melahirkan, kelompok hasil *pap smear test* normal dan abnormal memiliki rata-rata yang hampir sama, yaitu 25,6 tahun untuk kelompok hasil normal dan 26,2 tahun untuk kelompok hasil abnormal seperti yang ditunjukkan pada Tabel 4.2.

4.1.2 Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi

Karakteristik pasien *pap smear* di Rumah Sakit “X” Surabaya tahun 2010 berdasarkan faktor siklus menstruasi diberikan oleh Tabel 4.3.

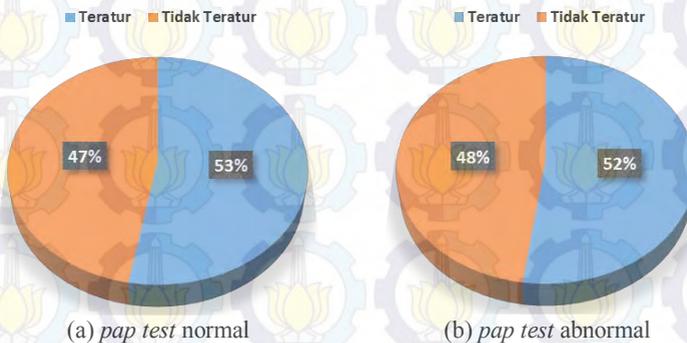
Tabel 4.3 Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi

		Hasil <i>Pap Smear</i>		Total
		Abnormal	Normal	
Siklus Menstruasi	Tidak Teratur	554	1144	1698
	Teratur	606	1282	1888
Total		1160	2426	3586

Pada kasus ini, dari 3586 pasien yang melakukan *pap smear test*, 1888 di antaranya memiliki siklus menstruasi teratur. Sementara sisanya, yaitu sebesar 1698 pasien, memiliki siklus menstruasi tidak teratur. *Odds ratio* yang diperoleh berdasarkan Tabel 4.3 adalah

$$OR = \frac{554 \times 1282}{1144 \times 606} = 1,024.$$

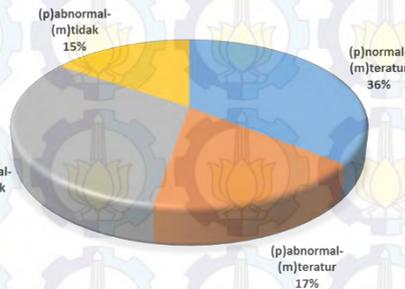
Artinya, pasien yang mengalami menstruasi tidak teratur memiliki risiko 1,024 kali lipat lebih besar untuk mendapatkan hasil *pap smear test* abnormal daripada pasien yang menstruasinya teratur.



Gambar 4.1 Karakteristik Pasien Berdasarkan Faktor Siklus Menstruasi Pada Masing-Masing Hasil *Pap Smear Test*

Seperti yang ditunjukkan pada Gambar 4.1(a), dari 2426 pasien dengan hasil *pap smear test* normal, 53% di antaranya memiliki siklus menstruasi teratur dan 47% sisanya memiliki siklus menstruasi tidak teratur. Sementara itu berdasarkan Tabel 4.3, pasien dengan hasil *pap smear test* abnormal yang berjumlah 1160 terdiri atas 606 pasien dengan siklus menstruasi teratur dan 554 pasien dengan siklus menstruasi tidak teratur. Gambaran secara visual dari kelompok *pap smear test* dengan hasil abnormal ini dapat dilihat pada Gambar 4.1(b) di mana persentase kejadian pasien yang memiliki siklus menstruasi teratur lebih besar daripada pasien yang siklus menstruasinya tidak teratur.

Secara keseluruhan, gambaran distribusi kejadian hasil *pap smear test* berdasarkan faktor siklus menstruasi dapat dilihat pada Gambar 4.2.



Gambar 4.2 Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Siklus Menstruasi

Berdasarkan Gambar 4.2, pasien dengan hasil *pap smear test* normal dan siklus menstruasi yang teratur memiliki persentase terbesar, yaitu sebesar 36% atau setara dengan 1282 kejadian dari total 3586 pasien. Sementara kejadian dengan persentase terkecil, yaitu 15%, merupakan karakteristik pasien dengan hasil *pap smear test* abnormal dan siklus menstruasi tidak teratur dengan jumlah kejadian sebesar 554.

4.1.3 Karakteristik Pasien Berdasarkan Faktor Paritas

Karakteristik pasien *pap smear* di Rumah Sakit “X” Surabaya tahun 2010 berdasarkan faktor paritas diberikan oleh Tabel 4.4.

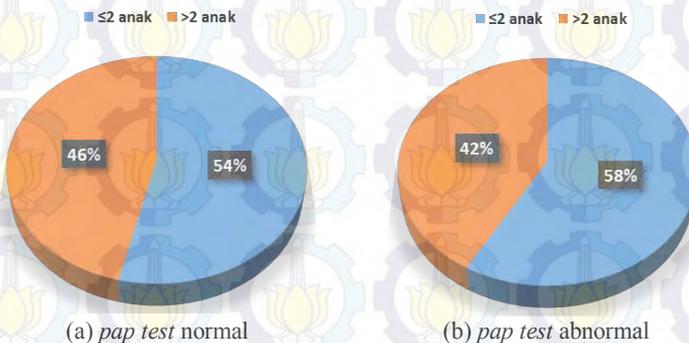
Tabel 4.4 Karakteristik Pasien Berdasarkan Faktor Paritas

		Hasil <i>Pap Smear</i>		Total
		Abnormal	Normal	
Paritas	>2 anak	484	1123	1607
	≤2 anak	676	1303	1979
Total		1160	2426	3586

Pada kasus ini, pasien yang pernah melahirkan maksimal dua kali berjumlah 1979 dan pasien yang pernah melahirkan lebih dari dua kali berjumlah 1607. Sehingga total keseluruhan pasien adalah 3586. Sementara itu, *odds ratio* yang didapatkan adalah sebesar

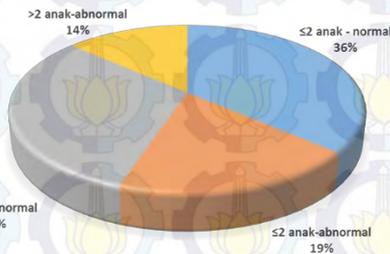
$$OR = \frac{484 \times 1303}{1123 \times 676} = 0,831.$$

Artinya, risiko pasien yang pernah melahirkan lebih dari dua kali untuk mendapatkan hasil *pap smear test* abnormal adalah sebesar 0,831 kali lebih besar dari pasien yang pernah melahirkan maksimal dua kali.



Gambar 4.3 Karakteristik Pasien Berdasarkan Faktor Paritas Pada Masing-Masing Hasil *Pap Smear Test*

Berdasarkan Tabel 4.4 dan Gambar 4.3, dari 2426 pasien dengan hasil *pap smear test* normal, 54% di antaranya atau setara dengan 1303 pasien pernah melahirkan maksimal dua kali. Sisanya, yaitu 1123 pasien, pernah melahirkan lebih dari dua kali. Sementara pada kelompok pasien dengan hasil *pap smear test* abnormal, pasien yang melahirkan maksimal dua kali memiliki persentase lebih besar dari pasien yang melahirkan lebih dari dua kali dengan jumlah kejadian masing-masing 676 dan 484.



Gambar 4.4 Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Paritas

Gambar 4.4 merupakan gambaran distribusi kejadian hasil *pap smear test* berdasarkan faktor paritas. Dapat dilihat pada Gambar 4.4, secara umum persentase kejadian pada kelompok *pap smear* dengan hasil normal memiliki total persentase yang lebih besar daripada kelompok *pap smear* dengan hasil abnormal, yaitu sekitar 67% dari total 3586 kejadian. Pada kelompok *pap smear* dengan hasil normal, persentase pasien yang pernah melahirkan maksimal dua kali dan lebih dari dua kali berturut-turut adalah 36% dan 31%. Sementara pada kelompok *pap smear* dengan hasil abnormal, persentase pasien yang pernah melahirkan maksimal dua kali adalah sebesar 19% dan untuk persentase pasien yang melahirkan lebih dari dua kali adalah sebesar 14%.

4.1.4 Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi

Tabel 4.5 merupakan karakteristik pasien *pap smear* di Rumah Sakit “X” Surabaya tahun 2010 berdasarkan faktor penggunaan alat kontrasepsi.

Tabel 4.5 Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi

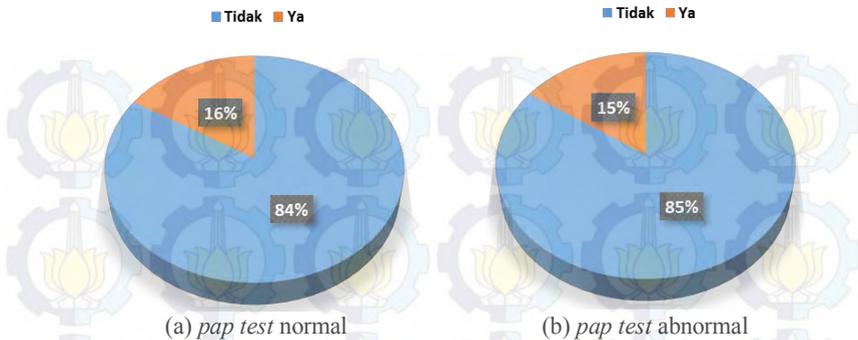
		Hasil <i>Pap Smear</i>		Total
		Abnormal	Normal	
Kontrasepsi	Ya	178	393	571
	Tidak	982	2033	3015
Total		1160	2426	3586

Pada kasus ini, jumlah pasien yang tidak menggunakan alat kontrasepsi sekitar 5 kali lebih banyak dari pasien yang menggunakan alat kontrasepsi dengan jumlah kejadian masing-masing 3015 dan 571. *Odd ratio* berdasarkan Tabel 4.5 adalah

$$OR = \frac{178 \times 2033}{393 \times 982} = 0,938.$$

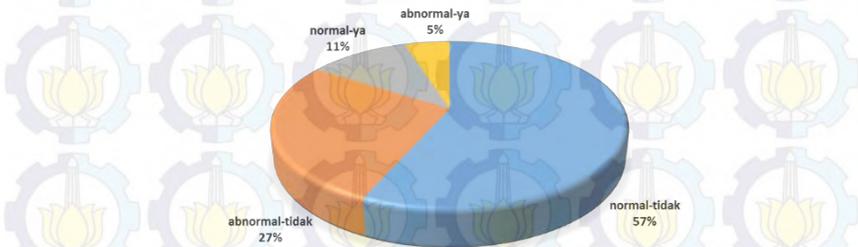
Artinya, pasien yang menggunakan alat kontrasepsi berisiko 0,938 lebih besar mendapatkan hasil *pap smear test* abnormal daripada pasien yang tidak menggunakan alat kontrasepsi.

Berdasarkan Tabel 4.5, jumlah pasien pada kelompok hasil *pap smear* normal yang tidak menggunakan kontrasepsi jauh lebih besar sekitar 5 kali daripada jumlah pasien yang menggunakan kontrasepsi. Hal ini juga dapat dilihat pada Gambar 4.5(a) di mana persentase pasien yang tidak menggunakan kontrasepsi sebesar 84%. Kondisi di mana jumlah pasien yang tidak menggunakan kontrasepsi lebih besar dari pasien yang menggunakan kontrasepsi juga berlaku pada kelompok hasil *pap smear* abnormal. Bila dilihat pada Gambar 4.5, persentase pasien yang tidak menggunakan kontrasepsi mencapai 85%; jauh lebih besar dari pasien yang menggunakan kontrasepsi dengan persentase hanya sebesar 15% atau setara dengan 178 dari 1160 kejadian.



Gambar 4.5 Karakteristik Pasien Berdasarkan Faktor Penggunaan Alat Kontrasepsi Pada Masing-Masing Hasil *Pap Smear Test*

Gambaran distribusi kejadian hasil *pap smear test* berdasarkan faktor penggunaan alat kontrasepsi ditampilkan pada Gambar 4.6.



Gambar 4.6 Karakteristik Pasien Pap Smear Test Berdasarkan Faktor Penggunaan Alat Kontrasepsi

Gambar 4.6 menunjukkan pasien dengan hasil *pap smear* normal namun tidak menggunakan alat kontrasepsi memiliki persentase terbesar, yaitu 57%. Disusul kemudian dengan kelompok pasien yang memiliki hasil *pap smear* abnormal dan juga tidak menggunakan alat kontrasepsi dengan persentase sebesar 27%. Sisanya, merupakan kelompok pasien dengan hasil *pap smear* normal dan menggunakan alat kontrasepsi serta

kelompok pasien dengan hasil *pap smear* abnormal namun menggunakan alat kontrasepsi yang masing-masing memiliki persentase sebesar 11% dan 5%.

4.1.5 Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran

Karakteristik pasien *pap smear* di Rumah Sakit “X” Surabaya tahun 2010 berdasarkan faktor riwayat keguguran disajikan pada Tabel 4.6.

Tabel 4.6 Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran

		Hasil <i>Pap Smear</i>		Total
		Abnormal	Normal	
Riwayat Keguguran	Pernah	809	1745	2554
	Tidak	351	681	1032
Total		1160	2426	3586

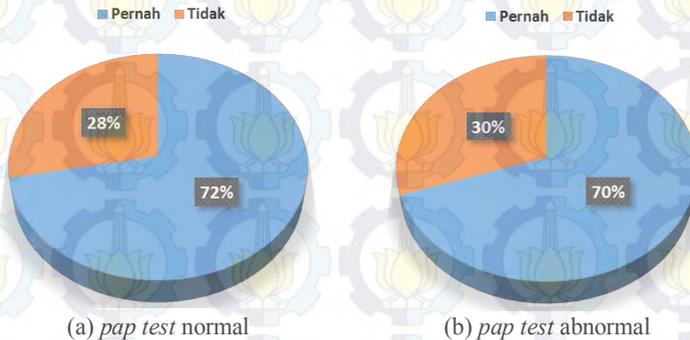
Berdasarkan Tabel 4.6, 2554 di antara 3586 pasien pernah mengalami keguguran sementara 1032 sisanya tidak pernah mengalami keguguran. *Odds ratio* yang diperoleh sebesar

$$OR = \frac{809 \times 681}{1745 \times 351} = 0,899.$$

Artinya, pasien yang pernah keguguran memiliki risiko 0,899 kali lebih besar mendapatkan hasil *pap smear test* abnormal dibandingkan dengan pasien yang tidak pernah mengalami keguguran.

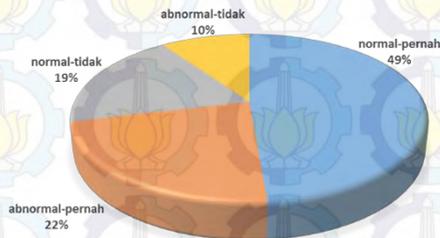
Tabel 4.6 menunjukkan bahwa dari 2426 pasien dengan hasil *pap smear* normal, pasien yang pernah mengalami keguguran memiliki jumlah sekitar 2,5 kali lebih besar bila dibandingkan dengan pasien yang tidak pernah mengalami keguguran. Hal ini juga terlihat dari Gambar 4.7(a) di mana persentase pasien yang pernah mengalami keguguran sebesar 72%, jauh lebih besar dari pasien yang tidak pernah mengalami keguguran. Persentase yang lebih besar pada kelompok pasien yang pernah mengalami keguguran juga terdapat pada kelompok hasil *pap smear* abnormal. Bila dilihat pada Tabel 4.6 dan Gambar 4.7(b), 70% dari total kejadian hasil *pap smear* abnormal atau setara dengan 809 kejadian

merupakan kelompok pasien yang pernah mengalami keguguran. Sementara sisanya, yaitu 30% atau setara dengan 351 kejadian merupakan pasien dengan karakteristik tidak pernah mengalami keguguran.



Gambar 4.7 Karakteristik Pasien Berdasarkan Faktor Riwayat Keguguran Pada Masing-Masing Hasil *Pap Smear Test*

Gambar 4.8 merupakan gambaran distribusi kejadian hasil *pap smear test* berdasarkan faktor riwayat keguguran.



Gambar 4.8 Karakteristik Pasien *Pap Smear Test* Berdasarkan Faktor Riwayat Keguguran

Pada Gambar 4.8, persentase terbesar berada pada kelompok pasien dengan hasil *pap smear test normal* dan pernah mengalami keguguran, yaitu sebesar 49%. Disusul kemudian dengan kelompok hasil *pap smear test abnormal* dan pernah mengalami keguguran sampai dengan kelompok hasil *pap smear abnormal* dan

tidak pernah mengalami keguguran yang memiliki persentase kejadian terkecil, yaitu sebesar 10%.

4.2 Klasifikasi Hasil *Pap Smear Test* Menggunakan Metode PPSSVM

Ada kalanya pada permasalahan klasifikasi biner, proporsi antara kelas negatif dan positif jauh berbeda. Akibatnya, *classifier* yang terbentuk akan menghasilkan akurasi yang sangat baik terhadap kelas dengan proporsi dominan, namun sangat buruk pada kelas dengan proporsi yang sedikit. Data hasil *pap smear test* pada penelitian ini memiliki proporsi kelas abnormal dan normal yang tidak seimbang. Oleh karena itu, proses *balancing* data perlu dilakukan sebelum proses klasifikasi dilakukan. Pada tahun 2014, Sain (2013) telah melakukan proses *balancing* pada data hasil *pap smear test* di Rumah Sakit “X” Surabaya tahun 2010 sehingga klasifikasi pada penelitian ini menggunakan hasil *balancing* data dari penelitian tersebut.

Pada permasalahan klasifikasi biner menggunakan SSVM, tahapan awal yang seharusnya dilakukan sebelum membangun model adalah mengidentifikasi kondisi data set yang ada. Ketika data set tersebut termasuk *linearly case* maka SSVM yang digunakan adalah SSVM linier, sementara bila data set merupakan *nonlinear case* maka digunakan SSVM nonlinier. Bila data set tersebut hanya memiliki dua variabel prediktor maka identifikasi dapat dengan mudah dilakukan, yaitu menggunakan *scatter plot*. Identifikasi akan sulit dilakukan ketika data terdiri atas beberapa variabel prediktor. Sejauh ini belum terdapat metode yang dapat digunakan untuk mengidentifikasi kondisi data set dengan banyak variabel prediktor. Ketika dihadapkan permasalahan seperti ini, biasanya kedua algoritma linier dan nonlinier akan diterapkan pada data. Kemudian, metode dengan *error rate* terkecil akan dipilih sebagai metode terbaik untuk klasifikasi data tersebut.

Mengingat adanya keterbatasan waktu pada penelitian ini maka data hasil *pap smear test* di Rumah Sakit “X” Surabaya tahun 2010 diasumsikan sebagai *nonlinear case*. Hal ini didasarkan pada hasil penelitian yang digagas oleh Purnami, dkk. (2013) yang

menerapkan beberapa metode SSVM nonlinier pada data hasil *pap smear test* tersebut dan mendapatkan performasi klasifikasi yang sangat baik.

4.2.1 Seleksi Parameter

Seleksi parameter merupakan langkah awal dalam pembentukan model klasifikasi sehingga nantinya diperoleh parameter yang dapat meminimumkan *error* dan memaksimalkan performasi klasifikasi. Parameter yang diseleksi merupakan parameter dari fungsi kernel dan parameter penyeimbang dalam fungsi objektif SSVM, yaitu ν . Pada penelitian ini fungsi kernel gaussian yang memiliki parameter μ digunakan untuk pemetaan nonlinier karena memiliki waktu komputasi yang lebih cepat. Sementara metode yang digunakan untuk seleksi parameter adalah metode *Uniform Design* (UD).

Proses seleksi parameter pada metode SSVM, PPSSVM1, dan PPSSVM2 menggunakan *software* MATLAB R2011b 32-bit pada komputer dengan spesifikasi prosesor Intel(R) Core i7 3,2GHz dan RAM 9 GB. Hasil dan waktu komputasi yang dibutuhkan untuk seleksi parameter untuk model SSVM, PPSSVM1, dan PPSSVM2 adalah sebagai berikut.

Tabel 4.7 Hasil dan Waktu Komputasi Seleksi Parameter

Metode	Parameter		Waktu Komputasi			
	μ	ν	Hari	Jam	Menit	Detik
SSVM	0,2880	10	0	14	37	21
PPSSVM1	0,7897	0,3109	5	17	21	40
PPSSVM2	0,7897	0,3122	3	6	7	30

Terlihat pada Tabel 4.7 bahwa secara umum metode PPSSVM memakan waktu komputasi paling lama dibandingkan dengan metode SSVM, khususnya metode PPSSVM1 yang memiliki waktu komputasi terlama. Perbedaan waktu komputasi ini dipengaruhi oleh jumlah data yang cukup besar serta kerumitan fungsi objektif dari masing-masing metode.

4.2.2 Pembentukan Model Klasifikasi Hasil *Pap Smear Test*

Pada pembahasan ini, model klasifikasi yang dimaksud adalah *decision function* pada persamaan (2.38) yang berguna untuk mengklasifikasikan suatu amatan yang diketahui nilai-nilai prediktornya. Pada persamaan tersebut, vektor \mathbf{u} dan nilai γ yang optimum dicari menggunakan algoritma Newton Armijo. Hasil optimisasi dari \mathbf{u} dan nilai γ untuk metode SSVM, PPSSVM1, dan PPSSVM2 terdapat pada Lampiran 5.

Pada *decision function* yang terbentuk, matriks \mathbf{A} mewakili variabel prediktor dari data hasil kanker serviks di Rumah Sakit “X” Surabaya tahun 2010. Sementara matriks \mathbf{x} mewakili variabel prediktor dari data amatan baru yang ingin dicari kelas klasifikasinya.

A. Model SSVM

Decision function SSVM yang terbentuk adalah sebagai berikut.

$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x}))$$

dengan

$$g(\mathbf{x}) = K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma$$

$$= \left[\exp\left(-0,2880 \|\mathbf{x}^T - \mathbf{A}\|^2\right) \right] \begin{bmatrix} -0,0808 \\ -0,0806 \\ \vdots \\ 1,6723 \end{bmatrix} - 0,9112$$

B. Model PPSSVM1

Decision function PPSSVM1 yang terbentuk adalah sebagai berikut.

$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x}))$$

dengan

$$g(\mathbf{x}) = K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma$$

$$= \left[\exp\left(0,7897\|\mathbf{x}^T - \mathbf{A}\|^2\right) \right] \begin{bmatrix} -0,3796 \\ 0,0000 \\ \vdots \\ 1,1395 \end{bmatrix} - 0,5004$$

C. Model PPSSVM2

Decision function PPSSVM2 yang terbentuk adalah sebagai berikut.

$$f(\mathbf{x}) = \text{sign}(g(\mathbf{x}))$$

dengan

$$g(\mathbf{x}) = K(\mathbf{x}^T, \mathbf{A}^T) \mathbf{D} \mathbf{u} - \gamma$$

$$= \left[\exp\left(0,7897\|\mathbf{x}^T - \mathbf{A}\|^2\right) \right] \begin{bmatrix} -0,3687 \\ -0,3687 \\ \vdots \\ 1,1503 \end{bmatrix} - 0,5147$$

4.2.3 Evaluasi Klasifikasi Hasil *Pap Smear Test*

Evaluasi model klasifikasi merupakan tahapan yang penting dalam permasalahan klasifikasi. Pada tahapan ini, metode SSVM, PPSSVM1, dan PPSSVM2 diterapkan pada data hasil *pap smear test* Rumah Sakit “X” Surabaya tahun 2010. Penerapan ketiga metode ini akan menghasilkan kelas prediksi yang kemudian dibandingkan dengan kelas asli dari data. Dari sini, performansi klasifikasi metode SSVM, PPSSVM1, dan PPSSVM2 dapat dihitung.

Pada dasarnya, model klasifikasi di bidang medis digunakan untuk mendeteksi ada tidaknya penyakit dalam diri seorang pasien sehingga performansi klasifikasi tidak hanya dilihat dari nilai akurasi, namun juga meliputi nilai sensitivitas dan spesifisitas. Kedua nilai ini sama-sama penting sehingga sebuah model klasifikasi yang bagus harus memiliki nilai sensitivitas dan spesifisitas yang tinggi (Zhu, et al., 2010)

Kemampuan model untuk menunjukkan hasil positif terhadap pasien yang benar-benar mengidap suatu penyakit diukur menggunakan sensitivitas, sementara kemampuan model untuk menunjukkan hasil negatif pada pasien yang memang tidak menderita suatu penyakit diukur menggunakan spesifisitas. Misal suatu model klasifikasi memiliki nilai sensitivitas 99%. Hal ini berarti bahwa ketika seorang pasien yang menderita suatu penyakit tertentu didiagnosis oleh tenaga medis, 99% hasil klasifikasi akan menunjukkan pasien tersebut positif mengidap penyakit. Sementara bila model klasifikasi memiliki spesifisitas sebesar 99% maka ketika pasien yang tidak menderita penyakit apapun didiagnosis oleh tenaga medis, terdapat peluang 99% bahwa hasil diagnosis tersebut akan negatif.

Pada penelitian ini, data hasil *pap smear* akan terlebih dahulu dibagi menjadi set *training* dan *testing* menggunakan *5-fold CV* sebelum menghitung performansi model klasifikasi. Nantinya, setiap *fold* yang terbentuk akan menjadi data *training* maupun *testing* secara bergiliran sehingga hasil evaluasi performansi model akan lebih baik karena memiliki variasi estimasi yang lebih kecil.

A. Metode SSVM

Tabel klasifikasi pada metode SSVM diberikan oleh Tabel 4.8. Penghitungan akurasi, sensitivitas, dan spesifisitas untuk *fold* 1 pada data *training* menggunakan SSVM adalah.

$$\text{Akurasi} = \left(\frac{1553+1589}{1553+1589+287+379} \right) \times 100\% = 82,51\%$$

$$\text{Sensitivitas} = \left(\frac{1589}{1589+287} \right) \times 100\% = 84,70\%$$

$$\text{Spesifisitas} = \left(\frac{1553}{1553+379} \right) \times 100\% = 80,38\%$$

Sementara pada data *testing* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{412+428}{412+428+40+70} \right) \times 100\% = 88,42\%$$

$$\text{Sensitivitas} = \left(\frac{428}{428+40} \right) \times 100\% = 91,45\%$$

$$\text{Spesifisitas} = \left(\frac{412}{412+70} \right) \times 100\% = 85,48\%$$

Penghitungan yang sama juga dilakukan terhadap *fold* 2 sampai dengan *fold* 5.

Tabel 4.8 Tabel Klasifikasi Metode SSVM

		<i>Fold</i>	Hasil Prediksi			
			<i>Training</i>		<i>Testing</i>	
			Positif	Negatif	Positif	Negatif
Kondisi yang Sebenarnya	Positif	1	1589	287	428	40
		2	1593	283	424	44
		3	1594	282	423	45
		4	1593	283	424	44
		5	1702	174	315	153
	Negatif	1	379	1553	70	412
		2	383	1549	66	416
		3	385	1547	64	418
		4	371	1561	78	404
		5	279	1653	170	312

B. Metode PPSSVM1

Tabel klasifikasi pada metode PPSSVM1 ditunjukkan oleh Tabel 4.9. Sementara penghitungan akurasi, sensitivitas, dan spesifisitas untuk *fold* 1 pada data *training* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{1664+1689}{1664+1689+187+268} \right) \times 100\% = 88,05\%$$

$$\text{Sensitivitas} = \left(\frac{1689}{1689+187} \right) \times 100\% = 90,03\%$$

$$\text{Spesifisitas} = \left(\frac{1664}{1664 + 268} \right) \times 100\% = 86,13\%$$

Penghitungan akurasi, sensitivitas, dan spesifisitas untuk *fold* 1 pada data *testing* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{438 + 449}{438 + 449 + 19 + 44} \right) \times 100\% = 93,37\%$$

$$\text{Sensitivitas} = \left(\frac{449}{449 + 19} \right) \times 100\% = 95,94\%$$

$$\text{Spesifisitas} = \left(\frac{438}{438 + 44} \right) \times 100\% = 90,87\%$$

Penghitungan yang sama juga dilakukan terhadap *fold* 2 sampai dengan *fold* 5.

Tabel 4.9 Tabel Klasifikasi Metode PPSSVM1

		<i>Fold</i>	Hasil Prediksi			
			<i>Training</i>		<i>Testing</i>	
			Positif	Negatif	Positif	Negatif
Kondisi yang Sebenarnya	Positif	1	1689	187	449	19
		2	1697	179	441	27
		3	1685	191	453	15
		4	1686	190	452	16
		5	1799	77	420	48
	Negatif	1	268	1664	44	438
		2	272	1660	40	442
		3	276	1656	36	446
		4	277	1655	35	447
		5	156	1776	64	418

C. Metode PPSSVM2

Tabel klasifikasi pada metode PPSSVM2 diberikan oleh Tabel 4.10.

Tabel 4.10 Tabel Klasifikasi Metode PPSSVM2

		<i>Fold</i>	Hasil Prediksi			
			<i>Training</i>		<i>Testing</i>	
			Positif	Negatif	Positif	Negatif
Kondisi yang Sebenarnya	Positif	1	1671	205	437	31
		2	1663	213	445	23
		3	1660	216	448	20
		4	1659	217	449	19
		5	1783	93	435	33
	Negatif	1	244	1688	38	444
		2	248	1684	34	448
		3	247	1685	35	447
		4	229	1703	53	429
		5	140	1792	54	428

Penghitungan akurasi, sensitivitas, dan spesifisitas untuk *fold* 1 pada data *training* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{1688 + 1671}{1688 + 1671 + 205 + 244} \right) \times 100\% = 88,21\%$$

$$\text{Sensitivitas} = \left(\frac{1671}{1671 + 205} \right) \times 100\% = 89,07\%$$

$$\text{Spesifisitas} = \left(\frac{1688}{1688 + 244} \right) \times 100\% = 87,37\%$$

Sementara pada data *testing* adalah sebagai berikut.

$$\text{Akurasi} = \left(\frac{444 + 437}{444 + 437 + 31 + 38} \right) \times 100\% = 92,74\%$$

$$\text{Sensitivitas} = \left(\frac{437}{437 + 31} \right) \times 100\% = 93,38\%$$

$$\text{Spesifisitas} = \left(\frac{444}{444 + 38} \right) \times 100\% = 92,12\%$$

Penghitungan yang sama juga dilakukan terhadap *fold 2* sampai dengan *fold 5*.

Rangkuman hasil evaluasi performansi SSVM, PPSSVM1, dan PPSSVM2 pada permasalahan klasifikasi hasil *pap smear test* di Rumah Sakit “X” Surabaya Tahun 2010 disajikan dalam Tabel 4.11 untuk data *training* dan Tabel 4.12 untuk data *testing*.

Tabel 4.11 Evaluasi Performansi Model SSVM, PPSSVM1, dan PPSSVM2 (%) Pada Data *Training*

Evaluasi Performansi	<i>Fold</i>	SSVM	PPSSVM1	PPSSVM2
Akurasi	1	82,51	88,05	88,21
	2	82,51	88,16	87,89
	3	82,48	87,74	87,84
	4	82,83	87,74	88,29
	5	88,10	93,88	93,88
Rata-rata		83,69	89,11	89,22
Sensitivitas	1	84,70	90,03	89,07
	2	84,91	90,46	88,65
	3	84,97	89,82	88,49
	4	84,91	89,87	88,43
	5	90,72	95,90	95,04
Rata-rata		86,04	91,22	89,94
Spesifisitas	1	80,38	86,13	87,37
	2	80,18	85,92	87,16
	3	80,07	85,71	87,22
	4	80,80	85,66	88,15
	5	85,56	91,93	92,75
Rata-rata		81,40	87,07	88,53

Keterangan : Nilai-nilai yang dicetak tebal menunjukkan nilai tertinggi pada setiap kelompok data

Tabel 4.12 Evaluasi Performasi Model SSVM, PPSSVM1, dan PPSSVM2 (%)
Pada Data *Testing*

Evaluasi Performasi	<i>Fold</i>	SSVM	PPSSVM1	PPSSVM2
Akurasi	1	88,42	93,37	92,74
	2	88,42	92,95	94,00
	3	88,53	94,63	94,21
	4	87,16	94,63	92,42
	5	66,00	88,21	90,84
Rata-rata		83,71	92,76	92,84
Sensitivitas	1	91,45	95,94	93,38
	2	90,60	94,23	95,09
	3	90,38	96,79	95,73
	4	90,60	96,58	95,94
	5	67,31	89,74	92,95
Rata-rata		86,07	94,66	94,62
Spesifisitas	1	85,48	90,87	92,12
	2	86,31	91,70	92,95
	3	86,72	92,53	92,74
	4	83,82	92,74	89,00
	5	64,73	86,72	88,80
Rata-rata		81,41	90,91	91,12

Keterangan : Nilai-nilai yang dicetak tebal menunjukkan nilai tertinggi pada setiap kelompok data

Secara keseluruhan metode PPSSVM1 dan PPSSVM2 memiliki rata-rata performasi yang lebih baik dibandingkan dengan metode SSVM. Sebagaimana yang ditampilkan pada Tabel 4.11, PPSSVM1 memiliki rata-rata sensitivitas yang paling baik pada data *training* maupun *testing* dengan tingkat sensitivitas masing-masing sebesar 91,22% dan 94,66%. Sementara

PPSSVM2 memiliki rata-rata nilai akurasi dan spesifisitas tertinggi pada data *training* dan *testing*. Rata-rata nilai akurasi metode PPSSVM2 adalah sebesar 89,22% pada data *training* dan 92,84% pada data *testing*. Sedangkan untuk rata-rata nilai spesifisitas, pada data *training* PPSSVM2 memiliki nilai sebesar 88,53% dan 91,12% untuk data *testing*.

Selanjutnya dapat dilihat pada Tabel 4.11, *fold 5* memiliki nilai akurasi, sensitivitas, dan spesifisitas tertinggi pada data *training* baik dalam metode SSVM, PPSSVM1, dan PPSSVM2. Sementara pada data *testing*, *fold 3* menyumbang nilai sensitivitas terbesar pada metode PPSSVM1, yaitu sebesar 96,79%, dan nilai akurasi tertinggi pada metode PPSSVM2, yaitu 94,21%. Sedangkan pada metode PPSSVM2, *fold 2* memiliki nilai spesifisitas tertinggi, yaitu 92,95%.

4.2.4 Pemilihan Model Terbaik untuk Klasifikasi Hasil *Pap Smear*

Pemilihan model terbaik merupakan salah satu tahapan yang krusial dalam permasalahan klasifikasi di bidang medis. Pada penelitian ini, pemilihan model terbaik tidak terlepas dari pertimbangan nilai akurasi, sensitivitas, dan spesifisitas dari ketiga metode yang digunakan. Tingkat sensitivitas yang tinggi pada sebuah model klasifikasi secara khusus menjadi penting ketika dihadapkan pada permasalahan diagnosis penyakit yang sangat berbahaya. Sementara bila hasil diagnosis yang positif memerlukan *treatment* yang berbahaya atau biaya yang sangat mahal maka tingkat spesifisitas yang tinggi menjadi indikator krusial dalam sebuah model klasifikasi (American Association for Clinical Chemistry, 2014). Pada kasus klasifikasi hasil *pap smear* dalam penelitian ini, tingkat sensitivitas dan spesifisitas yang tinggi memiliki urgensi yang sama sehingga model klasifikasi yang nantinya terpilih harus sama-sama memiliki nilai sensitivitas dan spesifisitas yang paling optimum.

Kurva ROC (*Receiver Operating Characteristics*) adalah representasi grafis dari hubungan timbal balik antara sensitivitas dan spesifisitas (Erke & Pattynama, 1998). Penggunaan kurva

ROC pada bidang medis pertama kali dilakukan oleh Lusted pada tahun 1960an dan menjadi populer pada tahun 1970an. Saat ini, kurva ROC digunakan secara luas untuk evaluasi performansi dari sebuah model klasifikasi di bidang teknologi kesehatan (Goncalves, et al., 2014). Salah satu metode yang digunakan untuk menghitung performansi model klasifikasi pada kurva ROC adalah AUC (*Area Under The ROC Curve*). Obuchowski (2003) dalam Park (2004) menyatakan bahwa AUC adalah sebuah ukuran performansi dari model klasifikasi yang diinterpretasikan sebagai rata-rata nilai sensitivitas untuk semua kemungkinan nilai spesifisitas. Oleh karena itu, performansi beberapa model dapat dibandingkan dengan membandingkan nilai-nilai AUC-nya. Semakin tinggi nilai AUC maka semakin baik performansi dari model klasifikasi (Park, et al., 2004).

Pada penelitian ini, *fold* pada data *testing* yang memiliki nilai AUC tertinggi akan dipilih dari masing-masing metode. Kemudian, perbandingan performansi ketiga metode akan dilakukan dengan membandingkan nilai AUC dari masing-masing *fold* yang terpilih.

Kelas terbesar pada data penelitian ini adalah 1, sementara kelas terkecil adalah 0 sehingga dengan menggunakan persamaan (2.50), *cutoff* (c) yang terbentuk adalah

$$c_1 = 0 - 1 = -1$$

$$c_2 = \frac{0+1}{2} = 0,5$$

$$c_3 = 1+1 = 2.$$

Sementara itu, jumlah kejadian pada kelas 1 (n_1) dan kelas 0 (n_0) masing-masing adalah 468 dan 482. Berdasarkan *cutoff* dan jumlah kejadian pada masing-masing kelas ini, penghitungan nilai AUC dan pembentukan kurva ROC untuk masing-masing metode dapat dijabarkan sebagai berikut.

A. Metode SSVM

Jumlah kelas 1 ($Y \geq c$) pada kelas 1 dan kelas 0 di setiap *cutoff* diberikan oleh Tabel 4.13.

Tabel 4.13 $s_1(c)$ dan $s_0(c)$ Pada Masing-Masing Fold Metode SSVM

c		<i>Fold</i>				
		1	2	3	4	5
-1	$s_1(c)$	468	468	468	468	468
	$s_0(c)$	482	482	482	482	482
0,5	$s_1(c)$	428	424	423	424	315
	$s_0(c)$	70	66	64	78	170
2	$s_1(c)$	0	0	0	0	0
	$s_0(c)$	0	0	0	0	0

Koordinat kurva ROC yang merupakan nilai sensitivitas (TP) dan 1-spesifisitas (FP) kemudian dihitung dengan menggunakan hasil pada Tabel 4.13 yang dimasukkan ke dalam persamaan (2.51). Hasilnya ditampilkan oleh Tabel 4.14.

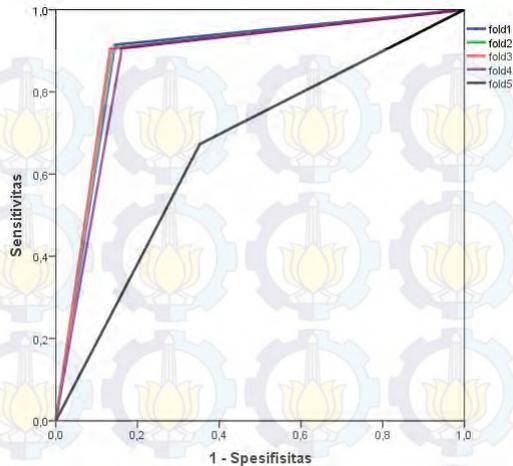
Tabel 4.14 Koordinat Kurva ROC Masing-Masing *Fold* SSVM

c		<i>Fold</i>				
		1	2	3	4	5
-1	TP	1,000	1,000	1,000	1,000	1,000
	FP	1,000	1,000	1,000	1,000	1,000
0,5	TP	0,915	0,906	0,904	0,906	0,673
	FP	0,145	0,137	0,133	0,162	0,353
2	TP	0,000	0,000	0,000	0,000	0,000
	FP	0,000	0,000	0,000	0,000	0,000

Berdasarkan koordinat kurva ROC pada Tabel 4.14 maka kurva ROC yang terbentuk dapat dilihat pada Gambar 4.9 dengan nilai AUC yang disajikan pada Tabel 4.15.

Tabel 4.15 Nilai AUC Pada Masing-Masing *Fold* SSVM

<i>Fold</i>				
1	2	3	4	5
0,885	0,885	0,886	0,872	0,660



Gambar 4.9 Perbandingan Kurva ROC dari Kelima *Fold* SSVM

Terlihat pada Tabel 4.15 bahwa *fold* ketiga memiliki nilai AUC tertinggi, yaitu sebesar 0,886. Oleh karena itu, *fold* 3 dipilih untuk mewakili metode SSVM untuk pemilihan model terbaik.

B. Metode PPSSVM1

Jumlah kelas 1 ($Y \geq c$) pada kelas 1 dan kelas 0 di setiap *cutoff* diberikan oleh Tabel 4.16.

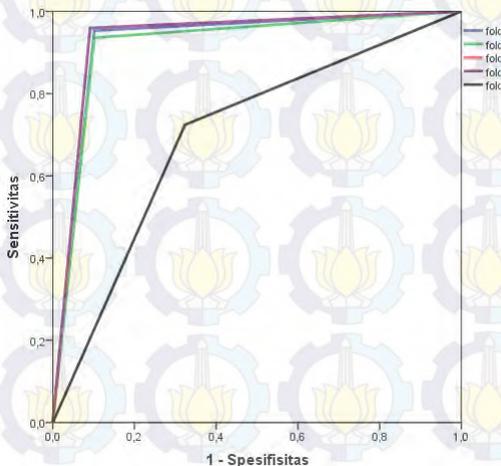
Tabel 4.16 $s_1(c)$ dan $s_0(c)$ Pada Masing-Masing *Fold* Metode PPSSVM1

c		<i>Fold</i>				
		1	2	3	4	5
-1	$s_1(c)$	468	468	468	468	468
	$s_0(c)$	482	482	482	482	482
0,5	$s_1(c)$	446	438	449	449	339
	$s_0(c)$	49	49	44	44	156
2	$s_1(c)$	0	0	0	0	0
	$s_0(c)$	0	0	0	0	0

Koordinat kurva ROC dihitung menggunakan persamaan (2.51) yang hasilnya dapat dilihat pada Tabel 4.17. Koordinat-koordinat ini yang membentuk kurva ROC kelima *fold* PPSSVM1 pada Gambar 4.10.

Tabel 4.17 Koordinat Kurva ROC Masing-Masing *Fold* PPSSVM1

<i>c</i>		<i>Fold</i>				
		1	2	3	4	5
-1	TP	1,000	1,000	1,000	1,000	1,000
	FP	1,000	1,000	1,000	1,000	1,000
0,5	TP	0,953	0,936	0,959	0,959	0,724
	FP	0,102	0,102	0,091	0,091	0,324
2	TP	0,000	0,000	0,000	0,000	0,000
	FP	0,000	0,000	0,000	0,000	0,000



Gambar 4.10 Perbandingan Kurva ROC dari Kelima *Fold* PPSSVM1

Sementara itu nilai AUC pada setiap *fold* dapat dilihat pada Tabel 4.18. Nampak bahwa *fold 3* dan *fold 4* memiliki nilai AUC tertinggi, yaitu sebesar 0,934. Oleh karena itu, *fold 3* akan mewakili

metode PPSSVM1 untuk pemilihan model klasifikasi hasil *pap smear test*.

Tabel 4.18 Nilai AUC Pada Masing-Masing *Fold* PPSSVM1

<i>Fold</i>				
1	2	3	4	5
0,926	0,917	0,934	0,934	0,700

C. Metode PPSSVM2

Jumlah kelas 1 ($Y \geq c$) pada kelas 1 dan kelas 0 di setiap *cutoff* diberikan oleh Tabel 4.19.

Tabel 4.19 $s_1(c)$ dan $s_0(c)$ Pada Masing-Masing *Fold* Metode PPSSVM2

<i>c</i>		<i>Fold</i>				
		1	2	3	4	5
-1	$s_1(c)$	468	468	468	468	468
	$s_0(c)$	482	482	482	482	482
0,5	$s_1(c)$	437	445	448	449	350
	$s_0(c)$	38	34	35	53	142
2	$s_1(c)$	0	0	0	0	0
	$s_0(c)$	0	0	0	0	0

Sementara koordinat kurva ROC dihitung menggunakan persamaan (2.51) yang hasilnya dapat dilihat pada Tabel 4.21. Koordinat-koordinat ini yang membentuk kurva ROC kelima *fold* PPSSVM1 pada Gambar 4.11.

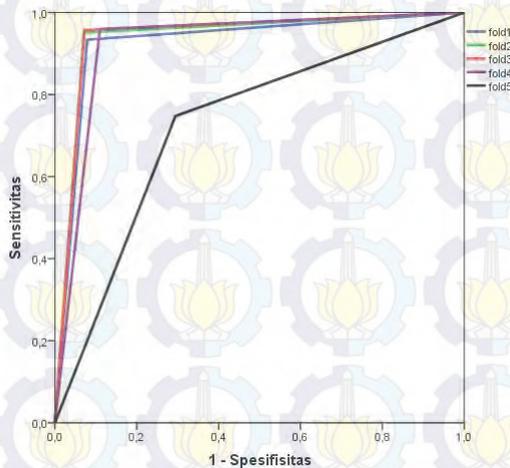
Nilai AUC pada setiap *fold* dapat dilihat pada Tabel 4.20. Terlihat bahwa nilai AUC terbesar terletak pada *fold* 3, yaitu sebesar 0,942.

Tabel 4.20 Nilai AUC Pada Masing-Masing *Fold* PPSSVM2

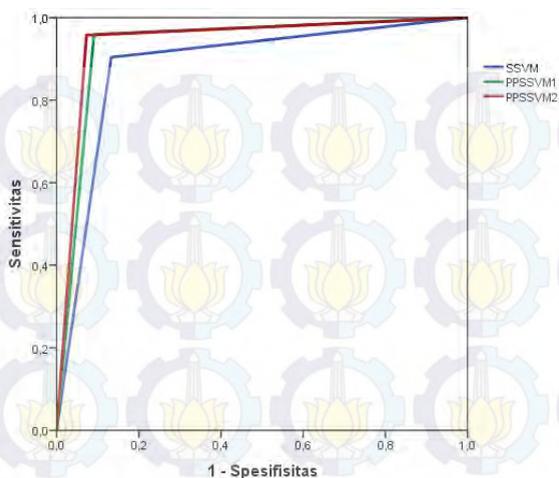
<i>Fold</i>				
1	2	3	4	5
0,927	0,940	0,942	0,925	0,727

Tabel 4.21 Koordinat Kurva ROC Masing-Masing *Fold* PPSSVM2

<i>c</i>		<i>Fold</i>				
		1	2	3	4	5
-1	TP	1,000	1,000	1,000	1,000	1,000
	FP	1,000	1,000	1,000	1,000	1,000
0,5	TP	0,934	0,951	0,957	0,959	0,748
	FP	0,079	0,071	0,073	0,110	0,295
2	TP	0,000	0,000	0,000	0,000	0,000
	FP	0,000	0,000	0,000	0,000	0,000

**Gambar 4.11** Perbandingan Kurva ROC dari Kelima *Fold* PPSSVM2

Pada pemilihan model terbaik untuk klasifikasi hasil *pap smear test* ini, baik metode SSVM, PPSSVM1, maupun PPSSVM2 sama-sama diwakili oleh *fold 3*. Kurva ROC dari *fold 3* pada setiap metode akan diplot dalam satu bidang *cartesius* untuk mempermudah analisis secara mendalam. Hasilnya ditampilkan pada Gambar 4.12.



Gambar 4.12 Kurva ROC untuk SSVM, PPSSVM1, dan PPSSVM2

Sementara rangkuman luas area di bawah kurva ROC (AUC) pada *fold 3* di setiap metode diberikan oleh Tabel 4.22.

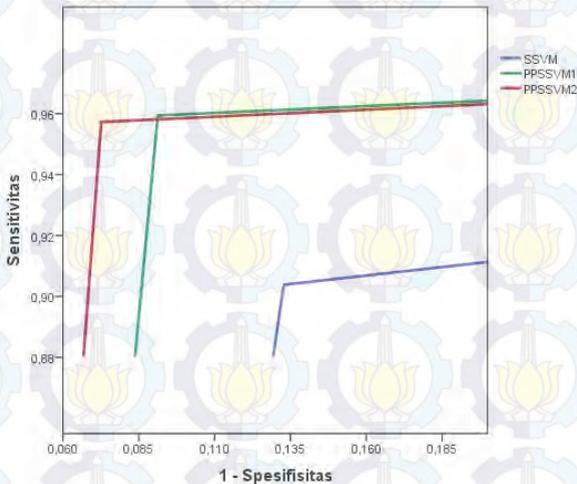
Tabel 4.22 Luas AUC dari Model SSVM, PPSSVM1, dan PPSSVM2

SSVM	PPSSVM1	PPSSVM2
0,886	0,934	0,942

Nampak pada Tabel 4.22 bahwa SSVM memiliki nilai AUC paling kecil bila dibandingkan dengan model PPSSVM. Berdasarkan nilai AUC yang dihasilkan, model PPSSVM1 dan PPSSVM2 masuk dalam kategori *excellent*, sementara model SSVM masuk dalam kategori *good*.

Sebagaimana yang disebutkan oleh Zweig dan Campbell (1993), semakin dekat plot ROC ke sudut kiri atas, semakin tinggi akurasi keseluruhan tes. Pada Gambar 4.12, dari ketiga kurva yang ada, model PPSSVM1 dan PPSSVM2 memiliki kurva ROC yang paling mendekati sudut kiri atas dibandingkan dengan model SSVM. Akibatnya, nilai AUC pada metode SSVM paling kecil dibandingkan dengan metode PPSSVM1 dan PPSSVM2. Sementara itu, dapat dilihat pada Gambar 4.13 bahwa kurva ROC

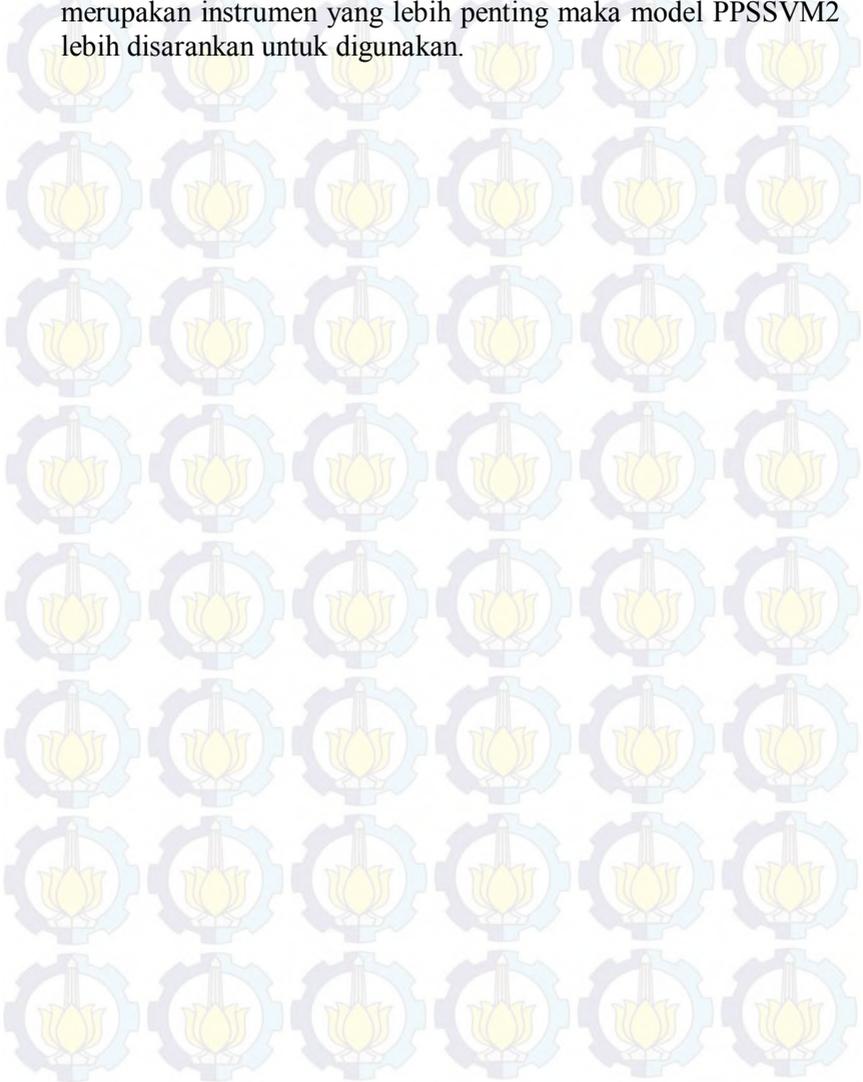
dari PPSSVM1 dan PPSSVM2 terlihat saling berhimpitan. Bila dilihat dengan seksama, bagian atas kurva ROC dari PPSSVM1 lebih mendekati sudut atas dari bidang plot dibandingkan bagian atas kurva ROC dari PPSSVM2. Hal ini menunjukkan bahwa PPSSVM1 memiliki nilai sensitivitas yang lebih tinggi dari PPSSVM2. Meskipun demikian, secara keseluruhan nilai AUC dari PPSSVM2 lebih besar dari PPSSVM1 meski hanya dengan selisih 0,008 ($0,942-0,934=0,008$).



Gambar 4.13 Kurva ROC untuk SSVM, PPSSVM1, dan PPSSVM2 (Diperbesar)

Berdasarkan analisis kurva ROC dan nilai AUC, dapat dikatakan bahwa kedua metode PPSSVM sama-sama baik dalam kasus pengklasifikasian hasil *pap smear test* di Rumah Sakit “X” Surabaya Tahun 2010. Metode PPSSVM2 mungkin memiliki nilai AUC yang lebih besar dari PPSSVM1. Namun, karena selisih dari kedua nilai AUC tersebut sangat kecil maka hal ini bisa diabaikan. Kedepannya, penggunaan model klasifikasi yang dibangun dengan metode PPSSVM1 dan PPSSVM2 dapat disesuaikan dengan kondisi yang diinginkan. Ketika tingkat sensitivitas yang tinggi

lebih diutamakan maka model klasifikasi PPSSVM1 disarankan untuk digunakan. Sementara bila tingkat spesifisitas yang tinggi merupakan instrumen yang lebih penting maka model PPSSVM2 lebih disarankan untuk digunakan.





Halaman ini sengaja dikosongkan

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan maka diperoleh kesimpulan sebagai berikut.

- 1) Karakteristik pasien *pap smear test* di Rumah Sakit “X” Surabaya tahun 2010 adalah sebagai berikut.
 - Berdasarkan karakteristik umur, rata-rata pasien berusia 43 tahun pada kelompok hasil normal dan 40 tahun pada kelompok hasil abnormal.
 - Pasien mengalami menstruasi pertama kali rata-rata pada umur sekitar 12 tahun baik pada kelompok hasil normal maupun abnormal.
 - Pada kelompok hasil normal dan abnormal, rata-rata pasien melahirkan untuk pertama kali pada usia sekitar 26 tahun.
 - Pasien dengan hasil *pap smear* normal mengalami siklus menstruasi tidak teratur sebesar 47%, sementara pada kelompok hasil abnormal sebesar 48%.
 - Persentase pasien yang memiliki jumlah anak lebih dari dua pada kelompok hasil *pap smear test* normal dan abnormal berturut-turut adalah 46% dan 42%.
 - Persentase pasien yang tidak menggunakan alat kontrasepsi sama-sama tinggi pada kelompok hasil *pap smear test* normal maupun abnormal, yaitu 84% dan 85%.
 - Pasien yang pernah mengalami keguguran memiliki persentase sebesar 72% pada kelompok hasil *pap smear test* normal dan 70% pada kelompok hasil *pap smear test* abnormal.
- 2) Secara keseluruhan metode PPSSVM1 dan PPSSVM2 memiliki performansi yang lebih baik daripada metode SSVM. Metode PPSSVM1 memiliki rata-rata tingkat sensitivitas tertinggi, yaitu 91,22% pada data *training* dan

94,66% pada data *testing*. Sementara PPSSVM2 memiliki nilai akurasi dan spesifisitas tertinggi di mana nilai akurasi pada data *training* dan *testing* adalah sebesar 89,22% dan 92,84%; sedangkan untuk nilai spesifisitas adalah sebesar 88,53% dan 91,12% untuk data *training* dan *testing*.

- 3) Kedua metode PPSSVM sama-sama baik dalam kasus pengklasifikasian hasil *pap smear test* di Rumah Sakit “X” Surabaya Tahun 2010. Model PPSSVM1 lebih baik digunakan bila tingkat sensitivitas lebih diutamakan. Sementara bila tingkat spesifisitas lebih diutamakan maka model PPSSVM2 lebih baik untuk digunakan.

5.2 Saran

Melalui proses analisis yang telah dilakukan dan hasil pembahasan yang didapatkan maka terdapat beberapa saran yang dapat diberikan, yaitu.

- 1) Penambahan faktor risiko kanker serviks pada penelitian berikutnya dapat dipertimbangkan karena masih terdapat beberapa faktor risiko yang belum dimasukkan, seperti usia pertama kali berhubungan seksual, kebiasaan merokok, dan lain sebagainya.
- 2) Pembagian kategori pada variabel yang berskala nominal maupun ordinal sebaiknya lebih diperhatikan dan didiskusikan secara mendalam dengan ahli medis yang bersangkutan.
- 3) Program PPSSVM memiliki performansi yang sangat baik, namun sangat tidak efisien dalam waktu komputasi, terutama pada bagian seleksi parameter. Oleh karena itu, pada penelitian selanjutnya dapat dicoba beberapa alternatif metode seleksi parameter yang lebih efisien dalam waktu komputasi.
- 4) Permasalahan komputasi untuk menyelesaikan berbagai metode SVM dapat dikembangkan lebih lanjut pada *software* yang bersifat *open source* sehingga aplikasi dari berbagai metode SVM dapat digunakan secara lebih luas.

DAFTAR PUSTAKA

Abbas, A., 2003. *Beberapa Faktor Risiko Kanker Serviks di Perjan RSUP. dr. Wahidin Sudirohusoso Makassar Tahun 2012*, Makassar: s.n.

American Association for Clinical Chemistry, 2014. *American Association for Clinical Chemistry*. [Online] Available at: <http://labtestsonline.org/understanding/features/reliability/start/1> [Diakses 24 November 2014].

Anonim, 2014. *Ayahbunda*. [Online] Available at: <http://www.ayahbunda.co.id/Artikel/Gizi+dan+Kesehatan/seputar.tes.pap.smear/001/001/1686/1/4> [Diakses 20 November 2014].

Australian Department of Health, 2013. *National Cervical Screening Program*. [Online] Available at: <http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/papsmear#1> [Diakses 21 Agustus 2014].

Cancer Research UK, 2014. *Cervical Cancer Risk Factors*. [Online] Available at: <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/cervix/riskfactors/cervical-cancer-risk-factors>

Chou, S., Shan, J., Guo, Y. & Zhang, L., 2010. Automated Breast Cancer Detection and Classification Using Ultrasound Image : A Survey, Pattern Recognition. Volume 43, pp. 299-317.

Cleveland Clinic, 2010. *Cleveland Clinic*. [Online] Available at: my.clevelandclinic.org/anatomy/female_reproductive_system

m/menstruation/hic-abnormal-menstruation.aspx

[Diakses 8 September 2014].

- Colak, E. et al., 2012. Comparison of Semiparametric, Parametric, and Nonparametric ROC Analysis for Continuous Diagnostic Tests Using a Simulation Study and Acute Coronary Syndrome Data. *Computational and Mathematical Methods in Medicine*, Volume 2012, p. 7.
- Cortes, C. & Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 1 September, 20(3), pp. 273-297.
- Dalimartha, S., 2004. *Deteksi Dini Kanker dan Simplisia Antikanker*. Jakarta: Penebar Swadaya.
- Dinas Kesehatan, 2014. *Hilangkan Mitos Tentang Kanker*, Jakarta: Kementerian Kesehatan Republik Indonesia.
- Erke, A. R. V. & Pattynama, P. M. T., 1998. Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology. *European Journal of Radiology*, pp. 88-94.
- Fatmawati, R. T., 2012. *Hubungan Paritas dengan Kejadian Kanker Serviks di Ruang Merak RSUD dr. Soetomo Surabaya*, Surabaya: s.n.
- Globocan, 2012. *GLOBOCAN 2012 : Estimated Cancer Incidence, Mortality and Prevelence Worldwide in 2012*. [Online] Available at: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx
- Goncalves, L., Subtil, A., Oliveira, M. R. & Bermudez, P. D. Z., 2014. ROC Curve Estimation : An Overview. *Statistic Journal*, 12(1), pp. 1-20.
- Hsu, C. W., Chang, C. C. & Lin, C. J., 2008. *A practical guide to Support Vector Classification*, Taipe: s.n.
- Huang, C. M., Lee, Y. J., Lin, D. K. J. & Huang, S. Y., 2007. Model Selection for Support Vector Machines via Uniform Design. *Computational Statistics & Data Analysis*, 1 Februari, 52(1), pp. 335-346.
- IARC, 2014. List of Classifications by Cancer Sites with Sufficient or Limited Evidence in Humans. Volume 1.

- Intansari, I. A. S., 2012. *Klasifikasi Pasien Hasil Pap Smear Test Sebagai Pendeteksi Awal Upaya Penanganan Dini Pada Penyakit Kanker Serviks di RS. "X" Surabaya Dengan Metode Bagging Logistic Regression*, Surabaya: Institut Teknologi Sepuluh Nopember.
- Lee, Y. J. & Mangasarian, O. L., 2001. SSVM: A Smooth Support Vector Machine for Classification. *Computational Optimization and Applications*, Volume 20, pp. 5-22.
- Lee, Y.-J., Yeh, Y.-R. & Pao, H.-K., 2012. Introduction to Support Vector Machines and Their Applications in Bankruptcy Prognosis. Dalam: J. Duan, W. K. Hardle & J. E. Gentle, penyunt. *Handbook of Computational Finance*. Berlin: Springer Berlin Heidelberg, pp. 731-761.
- Luo, L., Peng, H. & Zhou, Q., 2006. *A Study on Piecewise Polynomial Smooth Approximation to The Plus Function*. Singapura, IEEE, pp. 1-6.
- Mangasarian, L. O., 1998. *Generalized Support Vector Machines*, Madison, Wisconsin: University of Wisconsin.
- National Cancer Institute, 2012. *What You Need To Know About Cervical Cancer*. s.l.:s.n.
- Nurwijaya, H., Andrijono & Suheimi, H., 2010. *Cegah dan Deteksi Kanker Serviks*. Jakarta: Elex Media.
- Obuchowski, N., 2003. Receiver Operating Characteristic Curves and Their Use in Radiology. *Radiology*, Volume 229, pp. 3-8.
- Park, S. H., Goo, J. M. & Jo, C., 2004. Receiver Operating Characteristic (ROC) Curve : Practical Review for Radiologists. *Korean J Radiol*, March, 5(1), pp. 11-18.
- Purnami, S. & Embong, A., 2008. *Feature Selection and Classification of Breast Cancer Diagnosis Based on SVM*. Kuala Lumpur, Malaysia, s.n.
- Purnami, S., Embong, A., Zain, J. & Rahayu, S., 2009a. *A Comparison of Smoothing Function in Smooth Support Vector Machine*. s.l., s.n.

- Purnami, S., Embong, A., Zain, J. & Rahayu, S., 2009b. *Application of Data Mining Technique Using Best Polynomial SMOoth Support Vector Machine in Breast Cancer Diagnosis*. Langkawi Kedah, Malaysia, s.n.
- Purnami, S. W. et al., 2013. *Early Detection of Cervical Cancer Using Spline Smooth Support Vector Machine*, Surabaya: LPPKM ITS.
- Sain, H., 2013. *Combine Sampling Support Vector Machine untuk Klasifikasi Data Imbalanced*, Surabaya: ITS Press.
- Santosa, B., 2007. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis/Studi*. 1 penyunt. Yogyakarta: Graha Ilmu.
- Schneider, J., 1997. *Carnegie Mellon University : School of Computer Science*. [Online] Available at: <http://www.cs.cmu.edu/~schneide/tut5/node42.html>
- SEER US National Cancer Institute, 2011. *National Cancer Institute*. [Online] Available at: http://seer.cancer.gov/archive/csr/1975_2009_pops09/ [Diakses 20 November 2014].
- Sjamsuddin, S., 2001. Pencegahan dan Deteksi Dini Kanker Serviks. *Cermin DUnia Kedokteran*, Issue 133, pp. 9-14.
- Tira, D. S., 2008. *Risiko Jumlah Perkawinan, Riwayat Abortus, dan Pemakaian Alat Kontrasepsi Hormonal Terhadap Kejadian Kanker Serviks di Rumah Sakit Pelamonia Makassar Tahun 2006-2007*, s.l.: s.n.
- Verplancke, T. et al., 2008. Support Vector Machine Versus Logistic Regression Modeling for Prediction of Hospital Mortality in Critically Ill Patients with Haematological Malignancies. *BMC Medical Informatics and Decision Making*, 5 Desember, Volume 8, p. 56.
- WHO, 2013. *Human Papillomavirus (HPV) and Cervical Cancer*, s.l.: World Health Organization (WHO).

- Wulandari, I., 2014. *Piecewise Polynomial Smooth Support Vector Machine Untuk Klasifikasi Desa Tertinggal di Provinsi Kalimantan Timur Tahun 2011*, Surabaya: ITS Press.
- Wu, Q. & Wang, W., 2013. Piecewise-Smooth Support Vector Machine for Clasification. *Hindawi Publishing Corporation Mathematical Problems in Engineering*.
- Yakasai, I., Ugwa, E. & Otubu, J., 2013. Gynecological Malignancies in Aminu Kano Teaching Hospital Kano : A 3 Years Review. *Nigerian Journal of Clinical Practice*, 2 Februari, 16(1), pp. 63-66.
- Yastuti, H., 2011. Bagging Multivariate Adaptive Regression Splines (MARS) untuk Klasifikasi Pasien Hasil Pap Test Kanker Serviks (Studi Kasus di RS "X" Surabaya).
- Yuan, Y., Fan, W. & Pu, D., 2007. Spline Function Smooth Support Vector Machine for Classification. *Journal of Industrial and Management Optimization (JIMO)*, Agustus, 3(3), pp. 529-542.
- Yuan, Y. & Huang, T., 2005. A Polynomial Smooth Support Vector Machine for Classification. *Advanced Data Mining and Applications*, 22-24 July, Volume 3584, pp. 157-164.
- Yuan, Y., Yan, J. & Xu, C., 2005. Polynomial Smooth Support Vector Machine (PSSVM). *Chinese Journal of Computers*, pp. 9-17.
- Zhu, W., Zeng, N. & Wang, N., 2010. *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Pratical SAS Implementations*. s.l., NESUG org..
- Zweig, M. H. & Campbell, G., 1993. Receiver Operating Characteristic (ROC) Plots : A Fundamental Evaluation Clinical Medicine. *Clinical Chemistry*, pp. 561-577.



Halaman ini sengaja dikosongkan

LAMPIRAN

Lampiran 1. Data Hasil *Pap Smear Test*

Y	X1	X2	X3	X4	X5	X6
-1	68	12	12	1	1	1
-1	56	12	12	2	2	1
-1	31	12	12	1	2	1
-1	39	11	12	2	2	1
-1	72	12	12	1	1	1
-1	40	13	13	1	1	1
-1	39	12	14	1	2	1
-1	35	15	15	2	2	1
-1	37	14	16	2	1	1
-1	36	14	16	2	2	1
-1	34	13	16	2	1	1
-1	59	12	16	1	2	2
-1	33	14	16	2	1	1
-1	33	14	16	2	1	1
-1	46	13	17	1	2	1
-1	42	13	17	2	2	1
-1	55	12	17	1	2	2
-1	55	16	17	2	2	1
-1	36	15	17	1	2	1
-1	65	13	17	1	2	1
-1	29	13	17	1	2	1
-1	55	12	17	1	2	1
...
1	43	12	40	2	1	1
1	43	13	40	2	1	1
1	46	11	40	2	1	1
1	54	13	42	2	2	1
1	46	13	45	1	1	2

Lampiran 2. *Syntax* MATLAB untuk Seleksi Parameter

```
function Result = hibiscus(Label, Inst, Method, Type, Command)
Params.Method = Method;
Params.Type = Type;

if(exist('Command'))
    [pInd, pVal] = stread(Command, '%s%f', 'delimiter', ' ');
    for i=1:length(pInd)
        if(strcmp(pInd{i}, '-r'))
            Params.Mode = 'reduced';
            Params.Ratio = pVal(i);
        elseif(strcmp(pInd{i}, '-v'))
            Params.CV = pVal(i);
        else
            error('undefined parameter: %s', pVal(i));
        end
    end
end

if(~isfield(Params, 'Method'))
    Params.Method = 'SSVM';
end
if(~isfield(Params, 'Type'))
    Params.Type = 'ud';
end
```

```
if(~isfield(Params, 'Mode'))
    Params.Mode = 'full';
end
if(~isfield(Params, 'CV'))
    Params.CV = 5;
end

Elapse = cputime;

RIndex = 1:length(Label);
if (strcmpi(Params.Mode, 'reduced') & strcmpi(Params.Method, 'ssvm'))
    [RAIndex, RBIndex]=srsplit('svm', Label, Params.Ratio, 1);
    RIndex = [ RAIndex RBIndex];
elseif(strcmpi(Params.Mode, 'reduced') & strcmpi(Params.Method, 'ssvr'))
    RIndex = srsplit('svr', Label, Params.Ratio, 1);
end

if(strcmpi(Type, 'UD'))
    [CRange, GRange] = detRange(Inst, Params);
    P=[]; TErr_All=[]; VErr_All=[];
    [P, TErr_All, VErr_All] = ud(P, TErr_All, VErr_All, [13, 9], Label, Inst,
CRange, GRange, 1, Params, RIndex);
    [VErr, ind] = min(VErr_All);
    TErr = TErr_All(ind);
    Best_C = 2^P(ind, 1);
    Best_Gamma = 2^P(ind, 2);
elseif(strcmpi(Type, 'DOE'))
```

```
error([Mode, ', Not supported yet!!']);
elseif(strcmpi(Type, 'GRID'))
    CRange = [-5, 15];
    GRange = [-15, 3];
    GridSize = 20;
    [CGrid, GGrid] = meshgrid(linspace(CRange(1), CRange(2), GridSize),
linspace(GRange(1), GRange(2), GridSize));
    [TErr_All, VErr_All] = GridExplore(Label, Inst, CGrid, GGrid, Params,
RIndex);
    [val, ind1] = min(VErr_All); [val, ind2] = min(val);
    TErr = TErr_All(ind1(ind2), ind2);
    VErr = VErr_All(ind1(ind2), ind2);
    Best_C = 2^CGrid(ind1(ind2), ind2);
    Best_Gamma = 2^GGrid(ind1(ind2), ind2);
    P = Permute(linspace(CRange(1), CRange(2), GridSize), linspace(GRange(1),
GRange(2), GridSize));
    else
        error([Type, ', unknown!!']);
    end
    Elapse = cputime-Elapse; % toc

    Result.TErr = TErr;
    Result.VErr = VErr;
    Result.Best_C = Best_C;
    Result.Best_Gamma = Best_Gamma;
    Result.Elapse = Elapse;
```

```
Result.Points = 2.^P;
if(isfield(Params, 'Ratio'))
    Result.Ratio = Params.Ratio;
end
clear functions;

%*****
function [C_Range, G_Range] = detRange(Inst, Params)
if(strcmpi(Params.Mode, 'full'))
    G_Range = detGamma(Inst);
elseif(strcmpi(Params.Mode, 'reduced'))
    C_Range = [log2(1E+0), log2(1E+6)];
    G_Range = detGamma(Inst);
else
    error([Mode, ', unknown!!']);
end
%*****

function G_Range = detGamma(Inst)
Rows = length(Inst(:, 1));
dist = Inst-ones(Rows, 1)*mean(Inst);
```

```

k = sum((dist.^2)');
k = k(k>0);
k = sqrt(min(k));

G_Range = [log2(log(0.999)/-k), log2(log(0.150)/-k)];
%*****

function [P, TErr_All, VErr_All] = ud(P, TErr_All, VErr_All, Stage, Label, Inst,
CRange, GRange, th, Params, RIndex)

UDTable_13 = [7, 7; 5, 4; 12, 3; 2, 11; 9, 10; 6, 13; 3, 2; 11, 12; 13, 8; 10,
5; 1, 6; 4, 9; 8, 1];
UDTable_9 = [5, 5; 1, 4; 7, 8; 2, 7; 3, 2; 9, 6; 8, 3; 6, 1; 4, 9];
UDTable_5 = [3, 3; 1, 2; 2, 5; 4, 1; 5, 4];
CTable = []; eval(['CTable = UDTable_', num2str(Stage(th)), '!']);
Tlen = Stage(th);
CLower=CRange(1);   CUpper=CRange(2);   CLen=CUpper-CLower;   GLower=GRange(1);
GUpper=GRange(2);   Glen=GUpper-GLower;   UD_LV   = [(CLen/(Tlen-1))*(CTable(:, 1)-
1)+CLower(1), (Glen/(Tlen-1))*(CTable(:, 2)-1)+GLower(1)];
if th > 1
    UD_LV(1, :) = [];
end
[TErr, VErr] = GridExplore(Label, Inst, UD_LV(:, 1), UD_LV(:, 2), Params,
RIndex);
TErr_All = [TErr_All; TErr]; VErr_All = [VErr_All; VErr]; P = [P; UD_LV];

```

```
if th < length(Stage)
    [val, ind] = min(VErr_All);
    CRange = [P(ind, 1)-CLen/4, P(ind, 1)+CLen/4];
    GRange = [P(ind, 2)-GLen/4, P(ind, 2)+GLen/4];
    [P, TErr_All, VErr_All] = ud(P, TErr_All, VErr_All, Stage, Label, Inst,
    CRange, GRange, th+1, Params, RIndex);
end

%*****
function [TErr, VErr] = GridExplore(Label, Inst, CGrid, GGrid, Params, RIndex)

persistent counter;

[m, n] = size(CGrid); TErr = zeros(m, n); VErr = zeros(m, n);

flag = 0;

for i=1:m
    for j=1:n
        if isempty(counter)
            counter=1;
        else
            counter=counter+1;
        end
    end
end
```

```
end
disp(['trying number: ', num2str(counter)]);

if(strcmpi(Params.Method, 'ssvm'))
    % SSVM
    if (flag ~= i)
        K = SVKernel_C('rbf', Inst, Inst(RIndex,:), 2^GGrid(i, j));
        flag =flag + 1;
    end
    Model = K_ssvm_train(Label, K, ['-s 0 -c ', num2str(2^CGrid(i, j)),
'-v ', num2str(Params.CV)]);

    TErr(i, j)=Model.Err.Training(1);
    VErr(i, j)=Model.Err.Validation(1);
elseif(strcmpi(Params.Method, 'ssvr'))
    % SSVR
    if (flag ~= i)
        K = SVKernel_C('rbf', Inst, Inst(RIndex,:), 2^GGrid(i, j));
        flag =flag + 1;
    end
    Model = K_ssvm_train(Label, K, ['-s 1 -c ', num2str(2^CGrid(i, j)),
'-v ', num2str(Params.CV)]);

    TErr(i, j)=Model.Err.Training(1);
    VErr(i, j)=Model.Err.Validation(1);
else
    error('undefined method: %s', Params.Method);
```

```
end
end
end
%*****
function X = Permute(A, B)
LenA = length(A); LenB=length(B);
X = zeros(LenA*LenB, 2);
for i=1:LenA
    for j=1:LenB
        X((i-1)*LenB+j, :)=[A(j), B(i)];
    end
end
end
```

Lampiran 3. Syntax MATLAB untuk Pembentukan Model SSVM, PPSSVM1, dan PPSSVM2

```
function model=ssvm_train(label, inst, strParam)
% setting up parameters
params.s=0; params.t=2; params.c=100; params.g=0.1; params.r=1;
params.v=1;params.d=2;params.b=0;params.m=1;params.e=0.1;params.i=0;
[pInd, pVal] = strread(strParam, '%s%f', 'delimiter', ' ');
for i=1:length(pInd)
    if(strcmp(pInd{i}, '-s'))
        % ssvm type;G 0-SSVM, 1-SSVR (default: 0)
        params.s=pVal(i);
    elseif(strcmp(pInd{i}, '-t'))
        % kernel type;G 0-linear, 1-polynomial, 2-radial basis (default: 2)
        params.t=pVal(i);
    elseif(strcmp(pInd{i}, '-c'))
        % cost (default: 100)
        params.c=pVal(i);
    elseif(strcmp(pInd{i}, '-g'))
        % gamma (default: 0.1)
        params.g=pVal(i);
    elseif(strcmp(pInd{i}, '-r'))
        params.r=pVal(i);
    elseif(strcmp(pInd{i}, '-v'))
        % n-fold cross validation mode
        params.v=pVal(i);
```

```
elseif(strcmp(pInd{i}, '-d'))
    % degree of polynomial (default: 2)
    params.d=pVal(i);
elseif(strcmp(pInd{i}, '-b'))
    % constant term of polynomial (default: 0)
    params.b=pVal(i);
elseif(strcmp(pInd{i}, '-m'))
    % scalar term of polynomial (default: 1)
    params.m=pVal(i);
elseif(strcmp(pInd{i}, '-e'))
    % epsilon-insensitive value (default: 0.1)
    params.e=pVal(i);
elseif(strcmp(pInd{i}, '-i'))
    % alternative initial point (default: 1)
    params.i=pVal(i);
else
    error('undefined parameter: %s', pVal(i));
end
end
% start training
if params.r < 1
    if (params.s==0)
        % SSVM
        [RAIndex, RBIndex]=srsplit('svm', label, params.r, 1);
        RIndex = [ RAIndex RBIndex];
    else
        % SSVR
```

100

```
RIndex = srsplit('svr', label, params.r, 1);  
end  
[K, flag] = build_ker(params, inst, inst(RIndex,:));  
else  
RIndex = 1 : length(inst(:,1));  
[K, flag] = build_ker(params, inst, inst);  
end  
strParam = [];  
model=K_ssvm_train(label, K, strParam, params);  
model.RS=inst(RIndex, :);  
model.Space = flag;
```

%

```
function [K, flag] = build_ker(params, u, v)  
flag = 'dual';  
if (params.t==2)  
p = [params.g];  
K = SVKernel_C('rbf', u, v, p);  
elseif (params.t==0)  
[m, n] = size(u);  
if ((n > m) || (params.r < 1))  
K = SVKernel_M('linear', u, v);  
disp(['Solving linear SSVM/SSVR in dual space'])  
else
```

```
    K = u;  
    flag = 'primal';  
end  
else  
    p = [params.m params.b params.d];  
    K = SVKernel_C('poly', u, v, p);  
end
```

Lampiran 4. *Syntax* MATLAB untuk Prediksi

```
function [PredictedLabel, ErrRate]=ssvm_predict(label, inst, model)
```

```
%=====
```

```
if (strcmp('primal',model.Space))
```

```
    K = inst;
```

```
else
```

```
    K = build_ker(model.params, inst, model.RS);
```

```
end
```

```
[PredictedLabel, ErrRate]=K_ssvm_predict(label, K, model);
```

```
%=====
```

```
function K = build_ker(params, u, v)
```

```
if (params.t==2)
```

```
    p = [params.g];
```

```
    K = SVKernel_C('rbf', u, v, p);
```

```
elseif (params.t==0)
```

```
    K = SVKernel_M('linear', u, v);
```

```
    %K = u;
```

```
else
```

```
    p = [params.m params.b params.d];
```

```
    K = SVKernel_C('poly', u, v, p);
```

```
end
```

Lampiran 5. Parameter u dan γ untuk Model SSVM, PPSSVM1, dan PPSSVM2

SSVM		PPSSVM1		PPSSVM2	
u	γ	u	γ	u	γ
-0,08075	0,9112	-0,37956	0,5004	-0,36869	0,5147
-0,08055		0,00000		-0,36869	
-1,04232		-0,59395		-0,48370	
-0,42306		-0,37867		-0,36857	
-0,00081		0,00000		-0,36869	
-1,12250		-0,61959		-0,61421	
-0,51637		-0,39244		-0,01537	
-0,14986		-0,36704		-0,35626	
-0,42381		-0,42009		-0,39419	
-0,55494		-0,34318		-0,33118	
-0,21414		-0,34397		-0,33458	
-0,00606		-0,37559		-0,36815	
-0,06881		-0,21808		-0,21185	
-0,06881		-0,21808		-0,21185	
-1,58774		-0,43873		-0,50161	
-0,68786		-0,43110		-0,27994	
-0,22397		-0,29698		-0,32152	
0,00229		-0,37736		-0,36656	
-0,36040		-0,43525		-0,36689	
-0,02991		-0,32525		-0,28751	
-0,34936		-0,44600		-0,45900	
-0,36093		-0,28977		-0,09681	
-0,05087		-0,23427		-0,24220	
-0,03636		-0,14771		-0,28899	
....		
-0,12814		0,45721		0,30903	
0,75346		0,71127		0,68351	
0,82352		0,55384		0,81208	
1,67231		1,13946		1,15032	

Lampiran 6. Penghitungan Akurasi, Sensitivitas, dan Spesifisitas

A. Metode SSVM

	train_1	train_2	train_3	train_4	train_5	
TP	1589	1593	1594	1593	1702	
FN	287	283	282	283	174	
TN	1553	1549	1547	1561	1653	
FP	379	383	385	371	279	
total	3808	3808	3808	3808	3808	
TN+TP	3142	3142	3141	3154	3355	
TP+FN	1876	1876	1876	1876	1876	
TN+FP	1932	1932	1932	1932	1932	rerata
akurasi	82,51%	82,51%	82,48%	82,83%	88,10%	83,69%
sensi	84,70%	84,91%	84,97%	84,91%	90,72%	86,04%
spesi	80,38%	80,18%	80,07%	80,80%	85,56%	81,40%

	test_1	test_2	test_3	test_4	test_5	
	428	424	423	424	315	
TP	40	44	45	44	153	
FN	412	416	418	404	312	
TN	70	66	64	78	170	
FP	950	950	950	950	950	
total	840	840	841	828	627	
TN+TP	468	468	468	468	468	
TP+FN	482	482	482	482	482	rerata
TN+FP	88,42%	88,42%	88,53%	87,16%	66,00%	83,71%
akurasi	91,45%	90,60%	90,38%	90,60%	67,31%	86,07%
sensi	85,48%	86,31%	86,72%	83,82%	64,73%	81,41%
spesi	85,48%	86,31%	86,72%	83,82%	64,73%	81,41%

B. Metode PPSVM1

	train_1	train_2	train_3	train_4	train_5	
TP	1689	1697	1685	1686	1799	
FN	187	179	191	190	77	
TN	1664	1660	1656	1655	1776	
FP	268	272	276	277	156	
total	3808	3808	3808	3808	3808	
TN+TP	3353	3357	3341	3341	3575	
TP+FN	1876	1876	1876	1876	1876	
TN+FP	1932	1932	1932	1932	1932	rerata
akurasi	88,05%	88,16%	87,74%	87,74%	93,88%	89,11%
sensi	90,03%	90,46%	89,82%	89,87%	95,90%	91,22%
spesi	86,13%	85,92%	85,71%	85,66%	91,93%	87,07%

	test_1	test_2	test_3	test_4	test_5	
TP	449	441	453	452	420	
FN	19	27	15	16	48	
TN	438	442	446	447	418	
FP	44	40	36	35	64	
total	950	950	950	950	950	
TN+TP	887	883	899	899	838	
TP+FN	468	468	468	468	468	
TN+FP	482	482	482	482	482	rerata
akurasi	93,37%	92,95%	94,63%	94,63%	88,21%	92,76%
sensi	95,94%	94,23%	96,79%	96,58%	89,74%	94,66%
spesi	90,87%	91,70%	92,53%	92,74%	86,72%	90,91%

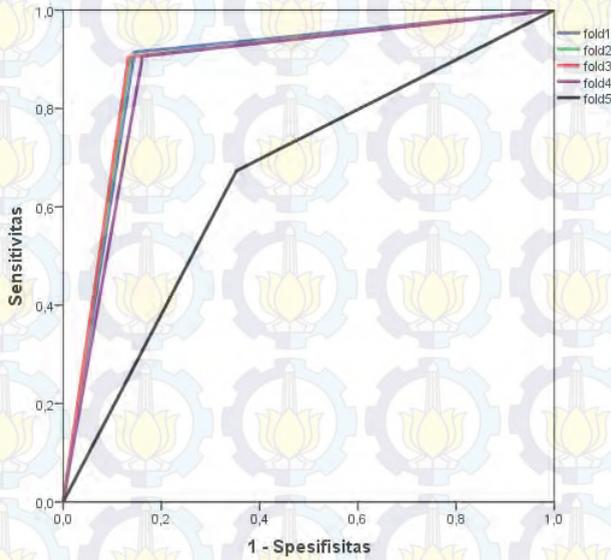
C. Metode PPSSVM2

	train_1	train_2	train_3	train_4	train_5	
TP	1671	1663	1660	1659	1783	
FN	205	213	216	217	93	
TN	1688	1684	1685	1703	1792	
FP	244	248	247	229	140	
total	3808	3808	3808	3808	3808	
TN+TP	3359	3347	3345	3362	3575	
TP+FN	1876	1876	1876	1876	1876	
TN+FP	1932	1932	1932	1932	1932	rerata
akurasi	88,21%	87,89%	87,84%	88,29%	93,88%	89,22%
sensi	89,07%	88,65%	88,49%	88,43%	95,04%	89,94%
spesi	87,37%	87,16%	87,22%	88,15%	92,75%	88,53%

	test_1	test_2	test_3	test_4	test_5	
TP	437	445	448	449	435	
FN	31	23	20	19	33	
TN	444	448	447	429	428	
FP	38	34	35	53	54	
total	950	950	950	950	950	
TN+TP	881	893	895	878	863	
TP+FN	468	468	468	468	468	
TN+FP	482	482	482	482	482	rerata
akurasi	92,74%	94,00%	94,21%	92,42%	90,84%	92,84%
sensi	93,38%	95,09%	95,73%	95,94%	92,95%	94,62%
spesi	92,12%	92,95%	92,74%	89,00%	88,80%	91,12%

Lampiran 7. Output SPSS Kurva ROC dan AUC

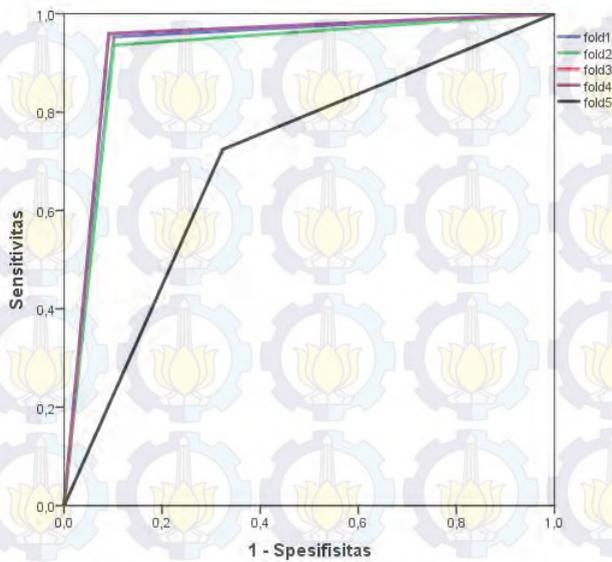
A. Metode SSVM



Area Under the Curve

Test Result Variable(s)	Area
fold1	,885
fold2	,885
fold3	,886
fold4	,872
fold5	,660

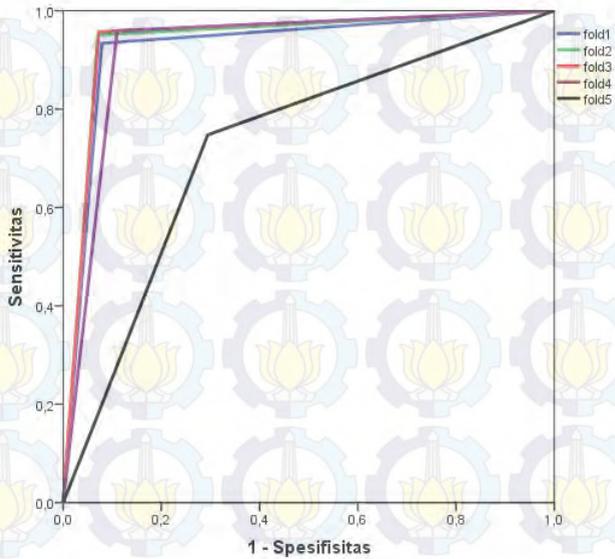
B. Metode PPSSVM1



Area Under the Curve

Test Result Variable(s)	Area
fold1	,926
fold2	,917
fold3	,934
fold4	,934
fold5	,700

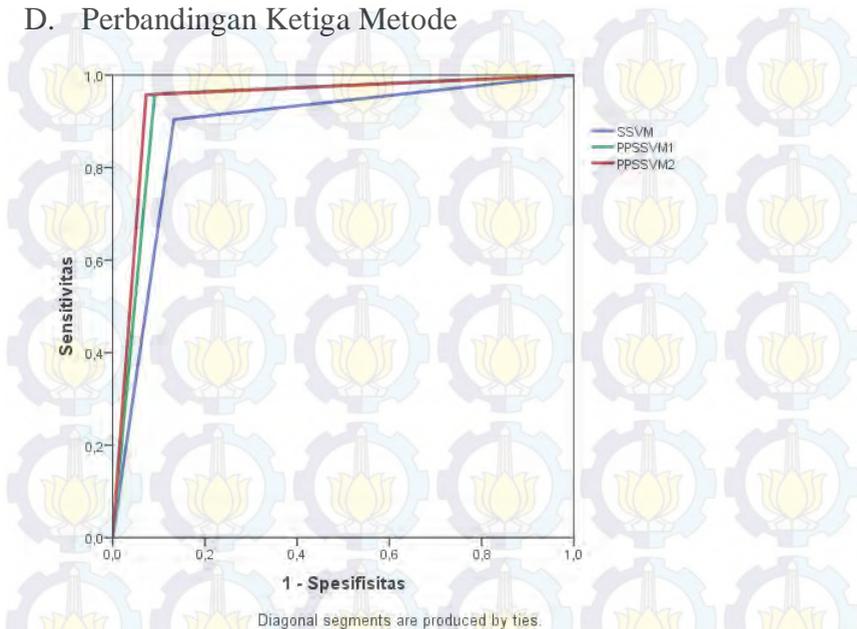
C. Metode PPSSVM2



Area Under the Curve

Test Result Variable(s)	Area
fold1	,927
fold2	,940
fold3	,942
fold4	,925
fold5	,727

D. Perbandingan Ketiga Metode



Area Under the Curve

Test Result Variable(s)	Area
SSVM	,886
PPSSVM1	,934
PPSSVM2	,942